781
672 H

THE VALUE OF PRAISE AND REPROOF AS INCENTIVES FOR CHILDREN

BY
ELIZABETH B. HURLOCK, M.A.

quir Fu an he Degree of Doctor
the De of Philosophy,
Philos ia niversit
Columbia

REPRINTED FROM
ARCHIVES OF PSYCHOLOGY
R. S. WOODWORTH, Editor
No. 71

STANFORD UNIVERSITY LIBRARIES
WITHDRAWN

NEW YORK
July, 1924

781
72 H

THE VALUE OF PRAISE AND REPROOF AS INCENTIVES FOR CHILDREN

BY
ELIZABETH B. HURLOCK, M.A.

Submitted in Partial Fulfillment of the [illegible] Doctor
of Philosophy, in the Department of [illegible]hy,
[illegible] the Degree of D[illegible]
in the Faculty [illegible]
[illegible]niversity

REPRINTED FROM
ARCHIVES OF PSYCHOLOGY
R. S. WOODWORTH, Editor

No. 71

NEW YORK
July, 1924

ACKNOWLEDGMENTS

The writer wishes to take this opportunity of expressing her gratitude to the many persons to whom she is indebted for co-operation in the carrying out of the experiment here reported. To Mr. Eugene Nifenecker, Director of the Bureau of Statistics and Research of the New York City School System; to Miss Kornmann and Miss Dowd of P. S. 119 Manhattan; and to Mr. Krause and Mr. Gainey of P. S. 5 Manhattan, through whose assistance the subjects were secured, she wishes to acknowledge her appreciation for their many kindnesses. For aid in the statistical treatment of the data, she is deeply indebted to Dr. H. E. Garrett of Columbia University. She is very grateful to Professor Pintner of Teachers' College, Columbia University, for his many valuable suggestions in planning the experiment. Above all, she wishes to express her appreciation to Professor Woodworth and Professor Poffenberger of Columbia University for their assistance in planning the experiment and for their many helpful suggestions and criticisms of the work from its beginning to its completion.

TABLE OF CONTENTS.

INTRODUCTION 5

SECTION I. Historical Survey 7

SECTION II. Conditions of the Experiment 21

1. The Subjects 21
2. The Tests 22
3. Procedure 22

SECTION III. Results 26

1. Total Results 26
2. Grade Comparisons 32
3. Sex Comparisons 38
4. Levels of Intelligence 46
5. Teachers' Ratings 55
6. Race Comparisons 60
7. Questionnaires 67
8. Correlations Between First and Second Tests 72

SECTION IV. Summary and Conclusions 75

BIBLIOGRAPHY 78

Digitized by the Internet Archive
in 2025

https://archive.org/details/valueofpraisered0000eliz

INTRODUCTION

The question of Motivation has, for the last decade or so, been coming more and more to the attention of the psychologist. Accompanying this has been an interest in the problem of the relative values of the different incentives which serve to motivate action. Workers in the fields of Animal and Industrial Psychology have long since recognized the importance of introducing some form of incentive if the best results are to be obtained. Educators, on the other hand, have been slow in evaluating the effectiveness of the different incentives which they use in dealing with children, with the consequence that this task has, to a certain extent, been taken over by a small number of psychologists.

Of all the incentives used during the school or college career of any individual, without doubt those most frequently and universally employed are praise and reproof. At once the question arises: "Are they of equal value in bringing about the desired result, or is one superior to the other?" The general assumption is that praise is the more effective of the two; but at the same time, one is forced to admit that reproof is the more frequently used.

The present investigation was conducted with a view towards determining, through experimental analysis, just how effective these two forms of incentives are, and which one tends to bring about improved work on the part of the child. In attempting thus to evaluate the two, a greater practical value is obtained if the incentives used are typical of a true school situation. Throughout the experiment, every effort was made to carry out this assumption as carefully and as accurately as possible.

The present study differs in several respects from studies previously made in the same field. The subjects were all school children rather than under-graduate or graduate college students. The situation was a typical school situation, and not a laboratory one. The tests were given as a part of the school program, and were not taken as a personal favor to the examiner.

Of the many different questions which might arise in connection with such a study, an attempt has been made to throw some light on the following:

1. What is the value of praise as compared with that of reproof?

2. Are older children more subject to the influence of praise and reproof than are younger children?

3. Do boys tend to react in a different manner than do girls of the same ages?

4. Of the three levels of intelligence,—superior, average and inferior,—are children belonging to one level more influenced by praise and reproof than children of the other two levels?

5. Does a child, accustomed to being rated as superior by his teacher, work better after praise or reproof than a child accustomed to being rated as inferior?

6. Can the common assumption, that negroes are more influenced by praise than by reproof, be verified?

7. Is any information concerning the children's reactions to the situation to be gained by means of a questionnaire study?

8. Finally, are the relative positions of the children within each group more changed as a result of practice, praise or reproof?

A resumé of the experimental studies of motivation precedes the experiment. An attempt has been made to cover the entire field, with special emphasis on all studies relevant to the present experiment, and from which the writer has taken many valuable suggestions.

SECTION I

Historical Survey

The field of Animal Psychology offers many examples of the use of incentives, mainly of reward and punishment. In training animals to run mazes, to discriminate between stimuli, or to form certain habits, the psychologist has had to resort to some kind of incentive. Hunter[6] says that "in most instances, the fundamental motives employed are hunger and the avoidance of painful stimulation. The animals are fed just enough to keep them in good physical condition. Electric shocks usually serve for the punishment, and are given when the animal makes an erroneous response." Washburn[23], when discussing the relative values of reward and punishment for animal learning, concludes that "punishment appears to produce more rapid learning than reward, unless it is so severe that it attaches itself to the whole learning situation. Punishment and reward combined, give, probably, better results than either alone."

One of the most recent studies of the use of incentives for animal learning, is the work of Rietta Simmons[18], published in January, 1924. In this research, the experimenter arranged the conditions of the experiment in such a way that a comparison of different types of incentives could be made, and their relative effectiveness determined. Using groups of albino rats, Dr. Simmons trained them on a maze problem under different conditions of incentive. The incentives employed were: bread and milk; sunflower seed; escape from the maze; return to the home cage; sex; litter and bread and milk under special conditions. Bread and milk was used as a standard, with which all of the other incentives were compared. The results of this study showed that bread and milk, either singly or in combination with return home, sunflower seed, sex and litter were much more effective incentives than escape or return home. Commenting on these results, the writer says, "These general results are not surprising. The four effective incentives call out the motives of hunger, sex or mother impulse, which we have assumed to be strong motives."

In the field of Industrial Psychology, one again finds examples of the use of different incentives to produce increased activity. Only one example from this field will likewise be given. The best study of this sort, from the writer's point of view, is the account of an unpublished study made by Dr. H. D. Kitson and reported by Book and Norvell[2]. Parts of this study will be quoted here. "The daily output of 40 hand compositors at the Lake Side Press in Chicago was studied for 12 months, and it was found that under a specially arranged bonus system set up in that establishment, the efficiency of every man was increased, on the average, from 58% to 97% in 20 weeks. All were experienced printers when they entered the plant, and had worked at the trade from 1 to 27 years. The average experience of these men was 10.3 years. Notwithstanding this fact, they made a gain in their output in 5 months, under the stimulus of a bonus system, of 67% over and above the average score made at the end of the first week of service. One man, who had been working at the trade for 27 years made a gain of 142%: another, who had been working at the trade for 14 years, increased his rate of work 282%. The most experienced half of these workmen increased their output, on the average, 50%, while the half which had the least experience at the trade, (6.3 years) made an average gain per man of 75%. And, although they were all considerably below the minimum bonus point in output when they started to work in the plant, the study showed that under the influence of the bonus system and a knowledge of their daily scores, 21 out of the 40 men had exceeded the bonus point at the end of the fourth week. In 5 months, 35 out of 40 men had reached this point: 50% of the men had increased their rate far beyond the bonus point, and so were receiving the extra pay."

Turning now from the fields of Animal and Industrial Psychology to more theoretical studies carried out in the psychological laboratory, we find that one of the earliest researches of this sort was made by Judd in 1905[8]. His problem was to find out "what was the effect of practice in a case when the subject was ignorant of his habit and of its effect." As material for this experiment, Judd used a table divided into two parts, one of which was screened from vision. On the part of the table that was visible, and next to the screen, he placed in succession 9 papers, each containing the picture of an

angle, differing in size and direction of slope of the sides from that which preceded it. After seeing each angle, the subject was told to place a dot on the screened-off portion of the table, in what he considered to be the exact continuation of the line. Throughout the experiment, the subject was kept in complete ignorance of his results. Using only one subject the experiment was carried on for 10 days, and the whole series of 9 lines was repeated 20 times each day. Judd has summarized the results of this study as follows: "The striking fact that appears in the results of these 10 days is that practice brings little if any change. The first day and the last day differ from each other about as did the first and second. There is no motive for improvement. The subject cannot see his results and cannot judge of their success or unsuccess."

Spencer[19] in 1923 followed up the work of Judd, repeating it as nearly as possible in every respect. His criticism of Judd's mathematical treatment of his results led him to carry out the work further and to see "what would be the effect of ignorance of results during practice if the errors were treated purely as errors of amount without regard to the direction." Using 4 subjects for the experiment, he found that, in 3 out of 4 cases, there was an improvement, and concluded, therefore, that the slight improvement which was indicated by Judd's results represented an actual improvement which Judd's method of averaging had, to some extent, disguised.

Wright[21] in 1906 investigated the effects of incentives on work and fatigue. "Each subject worked under two different mental conditions or attitudes: the one, that of working as hard and as long as he could, but with no idea of securing any specified result; the other, that of doing a specified task as long as his strength endured. Under the first condition, all incentive, such as watching his progress, was denied the subject. Under the second, however, he was not only allowed to watch his progress, but in addition to this, part of the time he was stimulated to action by being told to count his strokes." The Cattell ergograph was used, and as incentives for the second part of the series, blocks were inserted under the carriage of the ergograph, to which the subjects were required to reach as many times as possible during a given time. The results of all 4 subjects showed a gain in the work done when incentives were present. This fact, Wright interpreted as being due to "the mental attitudes of the subjects under the different con-

ditions of the experiment. The fatigue accompanying the work was not so great when the subject was working under the stimulus of a definite aim, notwithstanding the fact that at the same time he produced an increase in the amount of his work."

Arps[1], also using the ergograph, carried out an investigation to determine whether relatively complete awareness of results was a more favorable condition to efficiency than was a partial awareness of results. For each of the three subjects, there was a period of work in which the results were known to the subject, and corresponding to this, a period in which the results were unknown. The results of the experiment showed that of the total number of cases, 38% showed a higher efficiency value as compared with 19% of the unknown work period. From this, Arps concludes that "if the conditions of the experiment secure an equitable distribution of such complex factors involved in each work period, as adaptation, ergographic defects, fatigue, and the like, we may then conclude that the increased efficiency of the known periods over the unknown is conditioned by the conscious factor of awareness of results. This means that the organism functions more effectively when the ergographic tracings are immediately present stimuli."

Johanson[7] in 1922 published the results of an experiment, the object of which was to determine the influence of incentive and punishment on reaction-time. Using sound stimuli, Johanson put each of his 3 subjects through 3 series of presentations. "The first series was carried out according to normal procedure, with no complicating factors or conditions present. In the second or incentive series, the subject was informed that the time he took to react would be presented to him before the next succeeding reaction occurred. Thus, the time was used as an incentive to condition the reaction. In the third or punishment series, he was told that he would receive an electric shock only when he was not reacting quickly enough, but he was not told his previous record." The results thus obtained showed that the average improvement for the three subjects was 6.0% for the incentive series, and 14.8% for the punishment series. In order to take account of the practice effect, the punishment was introduced at different times for the three subjects, at the beginning, the middle and the end of the series. As almost no differences were found,

it showed that practice was very nearly eliminated before obtaining any results. Johanson gives the following explanation of the results of the experiment and the improvement found. "One of the possible, and to the writer, the most probable explanations is that a state of much keener attention, produced by an incentive and by the expectant punishment, motivates reactions very much prompter than those normally obtained under quieter and more restful conditions. The fact that there is a motive present, stimulating the subject to react, works the same in reaction-time as it does in all other types of responses to stimuli. How far the improvement may be carried on, and to what limits the time of a reaction may be reduced, is difficult of determination. The results show that the average time of the punishment series is about as fast as the individual is able to react, and it seems that the fear of punishment acts as a very powerful facilitator. That a state of keener attention or concentration of the attention was involved, is attested to not only by the premature reactions, but also by the feelings of the subjects. Especially during the punishment series, did the subjects evidence the usual signs of close attention as well as copious perspiration and nervous tension. This fact was also brought out in the introspective reports of the subjects."

The practice experiment technique has been employed by several investigators to study the influence of encouragement upon the individual's performance. Kirby[6], using as subjects 1350 New York school children from the third and fourth year classes, carried out two experiments of this sort, one in addition, and one in division. The practice in addition was carried on for 75 minutes, and in division, for 60 minutes, but the time was differently distributed for the different classes. Kirby has described the manner in which the tests were conducted thus: "After a practice period was finished and pencils laid down, the children were asked about the number of problems they had worked. This was done because the children were eager to tell their own scores and to learn the scores of others. It acted, too, as an immediate reward, and so as an incentive.....Just before beginning the second and each following practice period, the exact score of each child in the preceding day's practice was read, both the number of columns worked and the number correct. High scores with few errors were commended. If some score showed that

perhaps a child was adding beyond his best speed, as indicated by the number of errors, he was advised, not in a fault-finding way, but for the good of his own record, to try for greater accuracy with the hope of improving his score.....Children were told that their individual improvement was to be measured and they were shown that no matter how high or low their score, their final standing would be determined by the amount of gain made. They were shown that this was not primarily a contest among the individuals of a class, but an effort on the part of each one to surpass his own previous record. The children were encouraged to compare their last record with their own previous ones, and at times the scores were read to them in such a way as to indicate the gains made. Curves also were used to show improvement in actual scores.

"The desire to surpass one's own previous record was the incentive most appealed to; still the spirit of rivalry was strong.....The average number of problems done correctly in the preceding practice period was also read to the class before beginning the day's work. This group comparison served as an additional incentive, but would not satisfy the children as a substitute for their individual scores."

The results obtained from this procedure showed that for the group as a whole, practice in addition produced a median gain of 48% and a median loss of 0.4%. In division, there was a median gain of 75% during the practice period. Commenting on these results, the author says: "At the end of a test, there were frequent expressions of regret that the time was up, but almost no indications of indifference or ennui. The children were deriving immediate satisfaction from winning in a game where each was striving to be a victor, not under the enervating condition that he try to surpass someone who is natively his superior, but under the energizing condition that he surpass his own previous day's performance. This factor provided incentive for the slow child who ordinarily realized the futility of striving to do as well as the exceptional members of the class; it provided an incentive also for those exceptional members to exert their best effort, and thus precluded the dawdling that many methods cultivate among the more gifted members of a class.....Much of the gain in efficiency must have been due to the children's knowing their previous success, which in turn induced a pleasurable emotional tone to inspire continued intense effort and strengthen

associations formed. The intense effort that was displayed, the improvement that was made, and the emotional attitude that was aroused, seem to be features that should recommend the practice experiment as a method for school work."

Chapman and Feder[3] made a study very similar to that of Kirby, but introduced a control group as a check on the results found. The authors measured the initial efficiency of the subjects,—36 boys and girls from grade 5A in the Cleveland Observation School,—in three tests. These tests were the Thorndike Simple Addition Test, the Woodworth-Wells Cancellation Test, and the Digit-Symbol Test. Two equal groups were then formed on the basis of the initial scores in the addition test. Both groups were given extended practice in these tests, 10 minutes for addition, 1 for cancellation, and 5 for substitution, each day for 10 days. One of these groups worked under the normal conditions of the class-room, while the other was motivated by external incentives.

The average results were compared for the first, fifth and tenth day's performances. In addition, the motivated group made average scores of 14.7, 18.6 and 20.2 while the scores for the non-motivated group were 14.0, 11.7 and 11.9. In substitution, the motivated group made scores of 81, 155 and 215, and the non-motivated group, 79, 130 and 181. For cancellation, the scores were, for the motivated group, 32.0, 46.8 and 55.2, and for the non-motivated group, 34.0, 47.9 and 51.2. Commenting on these results, the authors say: "The general trend of the results shows that the incentive exerted a considerable effect on the amount of the product. It is significant that it is only in the cancellation test that the non-motivated group improves to the same extent as the motivated. In interpreting this cancellation result, it should be remembered that the length of practice over the 10 periods was only 10 minutes. We should not, therefore, expect the novelty of the test to wear off. It was only necessary for the subjects to do maximum work on each occasion for a single minute. This, apparently, can be accomplished without external incentives. It is only when we come to the other tests, which require more lengthy periods of work, that we notice the difference in gain. In the substitution test, there is 5 minutes of novel work, in the addition test, 10 minutes of work, which is very similar to the ordinary school routine. In both of these, greater improve-

ment was found than in cancellation, with the maximum amount in addition."

Stone and Colvin[20] made the question of "How to Study" the chief centre of interest and motive in a college course in Educational Psychology, and checked up the results of the 45 students in the group with a control group. Silent reading and comprehension were the two subjects of study most emphasized, and were measured before and after an intensive series of practices, consisting of 30 hours of reading text books in psychology and literary magazines. The results, as measured by Stone's extension of the Monroe's Silent Reading Test, showed that their average score after practice and study was 74% gain over that of the average scores before practice and study, while their average score in comprehension showed an increase of 84%. Moreover, they gained nearly three times as much in speed and three and a half times as much in comprehension as the control group. Individual gains of over 190% in comprehension and 160% in speed were found. The greatest gains existed in the cases of students who had made the lowest initial scores. The authors attributed the remarkable gains thus found to the fact that the students were able to increase their power of concentration by following a definite program of study, and by exerting every effort to concentrate while they were studying. Thus, the factor of concentration seemed to be the most important item concerned in securing increased rate and comprehension.

Perhaps the latest study of this sort, using the practice method technique, is the very extensive experiment of Book and Norvell[2] on the "Will to Learn," published in 1922. Their purpose in making this study "was to subject to experimental test the suggestions of Thorndike, Meuman, Book and Bryan regarding the dynamic influence of effort in learning, and the helpful effect of interest in improvement as such, and if possible, to determine the kind and amount of influence which this group of factors may exert on one's ability to improve." For subjects, 124 college students were used, 48 of these being men and 76 women. The tests consisted of 4 simple learning tests, the first involving 75 practice periods of of 30 seconds each, the other three, 45 practices, two of which were 1 minute long and the last, 1½ minutes. For each of these experiments, the total group was divided into two sections, the one acting as a control for the first two-thirds of the practice,

and the other as a stimulus section; while for the last one-third of the time the procedure for the two sections was reversed.

The authors have summarized the results found for this study as follows:

1. The stimulus sections made more improvement with a given amount of practice than did the control groups. This was true for the group as a whole and for the individual records.
2. The stimulus group, which had been making rapid and continuous improvement, ceased suddenly to show such gains when the special group of incentives was removed.
3. The control group, on the other hand, which had been making little or no progress in the practice, began suddenly to improve, and continued to gain more rapidly and consistently when definitely interested in its improvement as such, and when made to feel that further improvement was possible.
4. Little or no differences were found in the results made by the men and the women of the groups.
5. Fewer mistakes were made by the members of the stimulus sections than by the members of the control. This fact showed that the same groups of individuals not only ceased to gain in their scores, but also became more careless about their work when not interested in their improvement as such. When, on the other hand, they were subjected to special groups of incentives, they began at once to increase their output and accuracy.

Turning now from the studies in which the practice experiment technique was used, we may consider several studies which are similar, in some respects, to the experiment to be reported in the following pages, and to which the writer is indebted for many helpful suggestions. The first of these studies was made by Gilchrist in 1916[5], to determine the extent to which the work of college students is affected by praise or reproof. The subjects were 50 members of a class in Educational Psychology in the George Peabody College for Teachers. The entire class, as a group, was given the Courtis English Test, form 4B, and was then divided into two groups, "without regard to how well the individual members of the class had done in the test." The two groups thus formed were then placed in separate rooms, and the same test was given again to these groups. (The author does not state how long a time elapsed before giving the second test.) Before giving the second test, Gilchrist reproved Group "A" by saying: "A

hasty examination of the papers in the test just given shows that the members of this group did not do as well in the test as the average 12-year-old child would do. I ask you to take the test again." Group "B" was praised in the following manner: "A hasty examination of the papers in the test just given shows that the members of this group did exceptionally well. I ask you to take the test again." The results showed that Group "A" (reproved) made no improvement in spite of the practice, the average score on the first test being 59.8 and on the second, 59.21. Group "B," on the other hand, improved by 79%, from a score of 72.42 to 129.5. Commenting on these results, Gilchrist said: "Certainly we must look for the cause of this difference in the differences in environment of the two groups. And, so far as we are aware, the sole factor in creating a difference in the environment was the comment offered by the experimenter prior to giving the tests the second time. In this case, then, the effect of praise and reproof was a change in efficiency of 79%. Group 'A' showed a noteworthy peculiarity. The members of the class who had done well in the first test did not do so well in the second. The explanation of this fact is this. Insofar as the Courtis English test may be considered a test of mentality, the 'good' students made the higher scores in the first test, while the poorer students, on the other hand, made the lower scores. The results of the second test then show the 'good' students to have been more influenced by the comment than were the 'poor' students. It may have been that the former were unaccustomed to disparaging remarks, whereas the latter were inured to them."

In 1923, G. S. Gates and L. Q. Rissland[4] performed an experiment very similar to that of Gilchrist. The subjects were 74 Barnard College students who were given, individually, 2 simple performance tests, the "3" hole test, for motor coordination, and the color naming test. After the first trial had been given for both tests, the experimenter made the following comment to each third person tested: "That is really splendid. Do you always make such good scores? In a curve of distribution, your score would be way up here (indicating a position at the top of the curve). Your score was so good that I wonder if you would mind repeating the test?" The next subject was discouraged by saying: "Oh dear, that is really a very poor score! I am afraid that you would fall at

the bottom of the curve of distribution, etc." To the third subject, no comment was made concerning her performance. She was simply asked to repeat the test. This procedure was used alternately for all the subjects, and no attempt was made to equate the groups on the basis of their initial ability in both or either one of the tests. The results found may be summarized as follows:

1. The difference in average improvement of the 3 groups was slight, and the S. D.'s were relatively large.

2. The percentage of students who improved, deteriorated or remained the same was determined for each of the three groups. These showed that encouragement was more effective than discouragement, and discouragement than mere repetition.

3. Considering the effect of incentives upon groups differing in initial ability, it was found in general that the poorer the initial performance, the more effective was encouragement, and the less so was discouragement.

4. The effect of encouragement and discouragement was not found to be the same for the same individual in both tests.

The authors conclude this work with the following statement: "The results of this study seem to show a very slight difference in average improvement or even in the percentage of individuals who improve in the three groups. In this, the facts found are similar to those observed in experiments on fatigue, lack of fresh air, sleep or food; the external factors seem to be of relatively little importance in determining the score. Such difference as there is seems to be in favor of encouragement and discouragement, rather than mere repetition. We might say, then, that it is better to make an encouraging than a discouraging remark: that it is relatively poor individuals who are more likely to be unfavorably affected by discouragement than are relatively proficient persons; and that the effect of these two incentives does not seem to be constant for the two tests given."

Knight and Remmers,[10] using as subjects 10 college freshmen who had been subjected to a strenuous hazing process during a week of probation as pledges of a fraternity, carried out an experiment to determine the effect of motivation on such a group of men students. This group was told by a member of the fraternity that they were to assist in a scientific

experiment, and, as scores on this test were to count in the final election of members, they were to push themselves to the limit to do their best. The tests were given late on Friday night, following the week of probation. The material used was a set of Thorndike addition sheets, and the length of time of the experiment was divided into 24 work periods with a rest period of 10 minutes at the end of the first 12. At the end of each 5 minute period, the subjects were required to grade, on a scale of "A" to "F," the feeling tone which had accompanied that period of work. As a control group, 54 Juniors, mostly women, were used. They were given a similar test, lasting only over 7 work periods of 5 minutes each, and no introspective ratings were asked for. The results showed that, for the motivated group, an average of 21.9 columns was added per person in each of the 5 minute periods, while for the Juniors, only an average of 11. "There is nothing," say the authors, "to warrant the assumption that the Freshmen as a group were more skilful at this sort of thing by reason of previous training. The differential of motivation and mental attitudes is the explanation of the difference of averages." When the percentage of correct work was calculated, it was found that the Freshmen added correctly 85% of the work attempted during the first 12 periods, and decreased only to 83% during the last 12. The general conclusion given is as follows: "The above data support the assertion; 'The difference between a genuine motivation, such as the Freshmen had, and the kind of motivation a college class contains, is a difference which not only offsets extreme fatigue, but further offsets Freshmen vs. Junior ability, and in addition, produces twice as much work per unit of time with equal accuracy."

An introspective study of the effects of different incentives used by teachers was reported in December, 1923, by Laird[11]. He had 92 students from a class in Educational Psychology in the University of Wyoming think back to their high school days, and recall as impartially as possible the effect that different incentives used by their teachers had had upon their work. These incentives were to refer only to academic work, and not to behavior. Of the 15 incentives listed, two have direct bearing on the study to be reported later. The first, reprimands before others, was found to be not very effective in causing the subjects to improve their work. Twenty-six per cent of the total number reported that they worked better,

23.5% of whom were men and 28.5% women; 41.2% of the men and 39.2% of the women, with a total of 40.2%, did the same, while 35.5% of the men and 32.2% of the women said they did worse. Commenting on these results, Laird said: "Considering the percentage of students who fall into the 'same' or 'worse' groups, it is doubtful if public reprimands are justified in an attempt to secure more or better work for high school students." When public praise was used, on the other hand, by the teachers, 78.6% of the men reported that they did better work, 21.4% that they worked the same, and none as having done worse: 73.3% of the women did better, 26.7% the same, and again, none did worse. Thus, according to Laird, "Public commendation is much more effective in securing serious work from high school pupils than any of the strictly negative incentives. Reprimands, ridicule, sarcasm, low grades and additional assignments reduce the seriousness of the student's work rather than increase it. On the other hand, all the positive incentives reported on, as a rule, tend to result in an increased fervor of work. Some of the most commonly used incentives are the very ones that produce the most unfavorable results. Their commonness is no excuse for their continued use."

A similar report, relating to the incentives used in college classes, was also made by Laird.[12] The same incentives were listed, and practically the same results were found. His general conclusion to this study was: "The college student responds with more earnest work to the so-called positive incentives, rather than to the negative incentives. There is but little difference in college students and high school students in their responses to the different incentives."

A summary of the results of previous studies is rather difficult, due to the diversity of methods employed, and of the conditions under which the experiments were performed. In all cases, however, remarkable increases in performance were found to accompany the introduction of some incentive. On the other hand, these improvements were not found to be so great when a control group was used as a means of determining how much improvement might be expected to result from practice alone.

In the few studies in which the relative values of rewards and punishments, as incentives for human beings, were compared, the reward, no matter what form it took, seemed to be

the more effective of the two. The one exception to this is to be found in Johanson's study, in which physical punishment in the form of an electric shock, proved to be more influential in decreasing the subject's reaction time than did the positive incentive of knowledge of results made.

SECTION II

Conditions of the Experiment

1. *The Subjects*

The subjects used in this investigation were from P. S. 5 and P. S. 119, Manhattan, New York City. These are experimental schools under the direction of the Board of Research and Statistics of the New York Public School system. Although technically the former is a boys' school and the latter a girls' school, in both schools, in the lower grades, boys and girls are to be found in the same classes. The two schools are situated within 12 blocks of each other and the children attending them are of the same social status. In each case, there are about 40% of negro children, more of whom, however, are in the lower grades than in the higher ones.

The children selected for this study were from the 8th, 5th and 3rd grades of both schools. In every case, care was taken to have as nearly a random sampling as possible. A certain percentage was taken from each of the grade sections, and in addition to these, a few from the ungraded classes, so as to have a small group of children of very inferior intelligence.

TABLE I.

Distribution of Subjects According to School Grades.

	8th Grade.		*5th Grade.*		*3rd Grade.*		*Combined.*	
	No.	*%*	*No.*	*%*	*No.*	*%*	*No.*	*%*
BOYS.								
White.	54	38.3	50	37.9	30	22.2	134	32.8
Negro.	21	14.9	22	16.6	33	24.5	76	18.6
TOTAL.	75	53.2	72	54.5	63	46.7	210	51.4
GIRLS.								
White.	53	37.6	40	30.3	30	22.2	123	30.2
Negro.	13	9.2	20	15.2	42	31.1	75	18.4
TOTAL.	66	46.8	60	45.5	72	53.3	198	48.6
COMBINED.	141		132		135		408	

		No.	*%*
Whites.	(Combined)	257	62.9
Negroes.	"	151	37.1

Table 1 gives the distribution of subjects according to grades. From this, it may be seen that of the 408 children, 210 or 51.4% were boys, and 198 or 48.6% were girls. In

the boys' group, 76 or 18.6% out of this 51.4% were negro children, and in the girls' group, 75 or 18.4% of all were negroes, making a total of 151 or 37.1% negroes and 257 or 62.9% white children.

2. *The Tests*

Of all the possible standardized tests which might have been used as material for measuring the improvement due to praise or reproof, Intelligence Tests were finally decided upon. There were several reasons for this decision. The children who were to act as subjects for the experiment had taken similar tests the year before, and they knew their promotion depended to a great extent upon the outcome of the testing. Duplicating a situation of this sort insured an attitude of serious co-operation on the part of the children. In order that some information might be gained in connection with the problem of the value of incentives for children of different levels of intelligence, it was necessary to have the I. Q. scores of the subjects. Finally, since there is a growing interest in the question of the constancy of the I. Q. under the influence of various external factors, it was hoped that, as an interesting by-product of this study, the relative effects of encouraging or discouraging comments upon the subject's performance in the tests might be determined.

The National Group Intelligence Tests, Scale B, Forms 1 and 2, were used for the 8th and 5th grade children. The 3rd grade children were given forms A and B of the Otis Intelligence Scale, Primary Examination. In addition to these tests, all of the children in the three grades filled out a questionnaire at the end of the second testing.

3. *Procedure*

In order to obtain comparable results from the two schools, a standardized procedure for all grades in both schools was adopted. This procedure may best be illustrated by taking as an example, the 8th grade in the boys' school.

Seventy-five boys, chosen at random from the different sections of the 8th grade, were sent to the school assembly room at one time. There, under the direction of the assistant principal and several of the grade teachers, the writer gave them Form 1 of the National Intelligence Test Scale, according

to the standardized test procedure. After exactly one week, during which time the tests were corrected and three equivalent groups formed on the basis of the I. Q.'s obtained from the first tests and the C. A.'s, the boys were re-tested with Form 2 of the National Test, not together, as before, but in three separate groups. The first or "Control group" was given the test as at the previous time—according to the standardized procedure. The examiner merely said: "Please take out your pencils. You are to have a test like the one you took last week." Following the test, each boy filled out a copy of the questionnaire prepared for the Control group.

The second or "Praised" group was tested next. As soon as the boys were assembled, but before the examination blanks were passed out, the examiner made the following speech to the group. (This was used for all the children in the Praised groups, regardless of grade.) "Before beginning the test, I have a few things that I would like to say to you. I suppose you are wondering why I am giving you a test like the one you took last week? Well, I have just finished correcting all the papers for your grade, and I have selected you from the whole group who took the tests last week because of the very excellent work you did in that test. You not only made the best marks in your grade, but you did better than most boys (or girls) in grade (mentioning a grade several years higher than the one present) do in this test.

"Let me say a few words about your papers. They were exceptionally neat, the answers were put in the right places, and although some of you did not answer all of the questions, what you did answer, you answered well, and this counts more than answering all and getting half of them wrong.

"To-day, I am going to give you a test like the one you had last week. I want you to try not only to break your own record, but also to make this group stand first in the school. and set a standard for the others who do not do so well.

"I know that you can all do even better than you did last time, so work carefully and try just as hard as you possibly can. Remember, you are not only trying to break your own records, but you are going to set a standard for the others in your grade."

The test was then given according to the standardized test procedure, and at its completion, each boy filled out a questionnaire prepared for that group.

The procedure used in the case of the third or "Reproved" group was identical with that used with the Praised group, except that the comment preceding the test was this:

"Before beginning the test, I have a few things that I would like to say to you. I suppose you are wondering why I am giving you a test like the one you took last week? Well, I have just finished correcting all the papers for your grade, and I have selected you from the whole group who took the test last week because of the very poor work you did in that test. You not only did very badly in the test, but you are way below the standard for boys (or girls) of your age and grade. As the other boys (or girls) in your class did so well in this test, I know it isn't because you have not been taught these things in school. You have had the same chances to learn them as the others, but you just haven't tried.

"Your papers were slovenly, careless, and you made mistakes that even a second grade child would know better than to make. You are not supposed to answer all of the question on every page, but what you do, you are supposed to do carefully and correctly. This will count more in the end than doing all and making a lot of careless mistakes.

"I don't know whether you always do as badly in your school work as you did in this test. You certainly did badly enough in this test to feel thoroughly ashamed of yourselves, not only for your own sakes, but for your class record. It seems too bad for a group of boys (or girls) like you to bring down the class standard and hold back the others who really tried to do good work.

"I feel that it is only fair to give you another chance. So, to-day you are to have a test like the one you took last week. I don't know if you can do any better than you did last time—in fact, I rather doubt of you can—but I am going to give you this chance."

Thus, we have in each grade, three groups equivalent in I. Q.'s as based on the first tests, in chronological age, and in the number of white and negro children composing each. To each of these groups, second tests, equal in difficulty to the first, were given, but under different conditions. In the first case, the group merely repeated the test; in the second, the group repeated the test, but after being praised for their excellent work in the previous test, and urged to break their former records. The third group, on the other hand was told

just before taking the second test, that they had all failed in the first test, and that their poor work was holding back the others in their class who worked hard and tried to learn.

Any differences found after such a procedure, we may then say with a certain degree of confidence, are due to practice, praise or reproof. Using the control group as a standard for the amount of practice which may be expected for each of the three equivalent groups, we can thus determine with a greater or less degree of exactness, the effect that praise and reproof, as described above, have had on these groups of children.

SECTION III

Results

1. *Total Results*

In equating the three experimental groups, to be known throughout the study as "Control," "Praised" and "Reproved," several considerations had to be taken into account. The groups not only had to be equal in the average and variability scores of the first Intelligence Test, but when pairing the subjects, an attempt was always made to pair those who were of approximately the same chronological age. Likewise, distributions had to be so arranged as to have an equal number of white and negro boys or girls in each of the three groups.

Table 2 shows the distribution of children in the experimental groups for each of the three grades.

TABLE 2.

Distribution of Children in the Three Experimental Groups.

Groups.	*Control.*				*Praised.*				*Reproved.*			
	Boys.		*Girls.*		*Boys.*		*Girls.*		*Boys.*		*Girls.*	
	Wh.	*N.*	*Wh.*	*N.*	*Wh.*	*N.*	*Wh.*	*N.*	*Wh.*	*N.*	*Wh.*	*N.*
8th Grade.	18	7	18	4	18	7	18	4	18	7	17	5
5th Grade.	17	7	13	7	17	7	13	7	16	8	14	6
3rd Grade.	10	11	10	14	10	11	10	14	10	11	10	14
TOTALS.	45	25	41	25	45	25	41	25	44	26	41	25

TABLE 3.

Chronological Ages of the Children in the Three Experimental Groups.

Average.	*Control Group.*			*Praised Group.*			*Reproved Group.*		
	Boys.	*Girls.*	*Combined.*	*Boys.*	*Girls.*	*Combined.*	*Boys.*	*Girls.*	*Combined.*
8th Grade.	13.62	14.37	13.97	13.74	14.22	13.97	13.66	14.27	13.95
5th Grade.	11.29	12.25	11.73	10.96	12.10	11.48	11.12	12.35	11.68
3rd Grade.	9.26	10.46	9.90	9.22	10.66	9.99	8.88	10.45	9.72
σ									
8th Grade.	.99	1.39	1.25	1.14	1.28	1.23	1.18	1.50	1.37
5th Grade.	1.15	1.69	1.49	.94	1.69	1.45	1.32	1.52	1.54
3rd Grade.	.92	1.30	1.29	.98	1.71	1.60	.79	1.58	1.34
$P.\ E._{av.}$									
8th Grade.	.13	.20	.12	.15	.18	.12	.16	.21	.13
5th Grade.	.16	.25	.15	.13	.25	.15	.18	.28	.16
3rd Grade.	.13	.18	.13	.14	.23	.16	.12	.22	.13

In Table 3 are given the average chronological ages of the children in each of the three experimental groups. A consideration of the contents of this table shows that the three groups

are practically similar in chronological ages. The slight differences which do occur are too small to be of any significant value.

A comparison of the three experimental groups was first made from the total results, when no account was taken of the grade, sex, race or age factors within each group. Table 4 contains the results of the first testing.

TABLE 4.

I. Q. SCORES. FIRST TEST. TOTAL RESULTS.

	Control Group.	*Praise Group.*	*Reproved Group.*
Average.	98.77	98.74	98.82
σ	17.63	18.55	17.59
P.E. av.	1.01	1.06	1.01

RELIABILITY OF DIFFERENCES.

Groups.	*Actual Difference.*	*P.E. Difference.*	*Difference in P.E. Units.*	*Chances that true diff. is above 0.*
Control-Praised.	.03	1.46	.02	5054 in 10,000.
Control-Reproved.	–.05	1.43	.03	5081 in 10,000.
Praised-Reproved.	–.08	1.46	.05	5135 in 10,000.

From this, it may be seen that the average I. Q. score for the Control Group was 98.77, for the Praised, 98.74; and for the Reproved, 98.82. The measure of varability used, in this case, the standard deviation, shows likewise a very small amount of difference from one group to the other.

The reliability of the differences between the groups was calculated from the formula: P. E. (true-obt. diff, A-B) = $\sqrt{\text{P. E.}^2 \text{ (true-obt. A)} + \text{P.E.}^2 \text{ (true-obt. B)}}$. The results of these comparisons show that the actual differences, as well as the differences expressed in terms of P. E. units, are too small to be significant. In each case, the chances of a true difference existing between the groups are about 50-50, or even. We may thus conclude that the groups, taken as a whole, are equivalent, as judged by their performance in the first test.

Turning now to a consideration of the percentage of children in each group who gained, remained the same, or lost, we find the results summarized in Table 5.

TABLE 5.

PERCENTAGE OF CHILDREN WHO GAINED, REMAINED THE SAME, OR LOST IN THE THREE GROUPS. TOTAL RESULTS.

	Control Group.	*Praised Group.*	*Reproved Group.*
% Gaining.	52.2	78.7	80.2
% Same.	10.3	3.7	6.6
% Losing.	37.5	17.6	13.2

From the above table, it may be seen that the percentage of children who improved after praise and reproof is higher for both of these groups than for the Control group. While only 52% of the subjects in the Control Group show any improvement in the second test over the first, 79% of the Praised and 80% of the Reproved Groups did better. The difference between the later two groups is too small to be considered, but the difference between them and the Control Group is great enough to be significant of a true difference.

In order to determine the actual amount of gain in each group, the differences in gains between the groups, and the reliability of these differences, the individual gains were treated separately, apart from the original I. Q. scores from which they were obtained. This method was adopted because the gains themselves were found to vary less from individual to individual than did the original scores. It was thus hoped that when relatively small groups of individuals were compared the differences which might exist would not be obscured by the large variability of the group.

In Table 6 are presented the results for the three groups when treated in terms of the gains made in the second test, over and above the first.

TABLE 6.

RESULTS EXPRESSED IN TERMS OF INDIVIDUAL GAINS.

TOTAL RESULTS.

	Control Group.	*Praised Group.*	*Reproved Group.*
Averages.	1.20	6.19	6.33
σ	6.38	7.17	7.03
P.E. av.	.37	.41	.40
Range.	18 to –25	31 to –9	32 to –14

RELIABILITY OF DIFFERENCES.

Groups.	*Actual Difference.*	*P.E. Difference.*	*Difference in P.E. Units.*	*Chances that true diff. is above 0.*
Praised-Control.	4.99	.55	9.07	10,000 in 10,000.
Reproved-Control.	5.13	.54	9.50	10,000 in 10,000.
Praised-Reproved.	–.14	.57	.25	5,670 in 10,000.

The results of the comparison furnish interesting and conclusive data. Considering first the range of individual gains, we find that the largest gain occurred in the Reproved Group (32), the next largest in the Praised (31), while the largest made in the Control Group was only 18. On the other hand the largest loss occurred in the Control Group (—25) the next largest in the Reproved (—14) and the smallest in the Praised (—9). The average difference between the Control and Praised

Groups is 4.99 I. Q. points, between the Control and Reproved, 5.13 points; and between the Praised and Reproved Groups, only .14 points, in favor of the latter. When the reliability of these differences was determined, it was found that there is absolute certainty that the true difference between the Control and Praised, and Control and Reproved Groups is greater than O, while the chances of a true difference of over O in the case of the Praised and Reproved Groups are only 57 to 43, or almost even.

Diagram I.

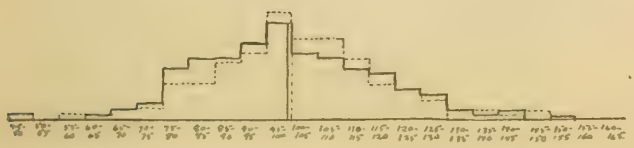
I. Q. Scores.

Total Results

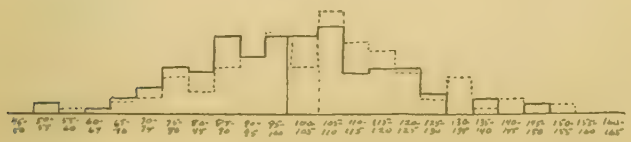
Control group.

— 1st Test.

---- 2nd Test.



Praised Group.

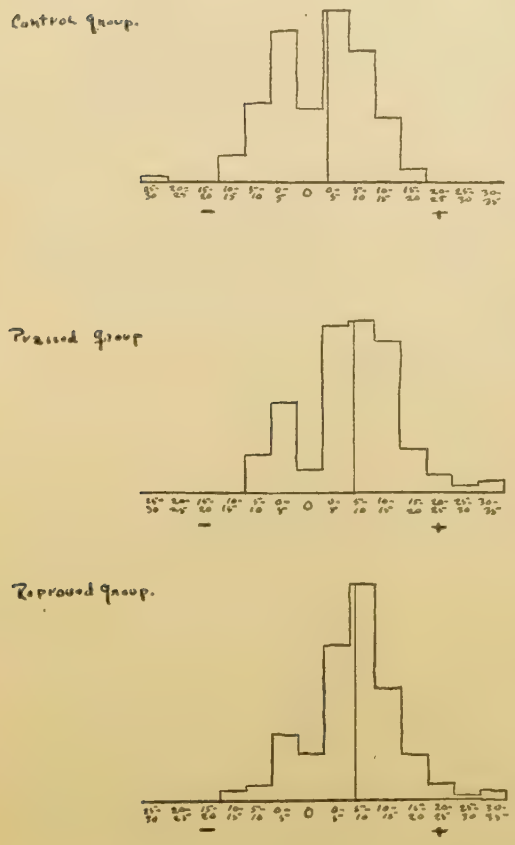


Reproved Group.

In Diagram 1 are shown the distributions of I. Q. scores for the three groups in both tests. The heavy black line represents the distribution for the first tests, and the dotted lines, the distributions for the second tests. The solid perpendicular line shows the average scores for the first tests, and the dotted one, the averages for the second tests. From these curves it may be seen that there is a tendency to shift scores to the upper end in both Praised and Reproved Groups, while

Diagram II.

Distribution of Gains.
Total Results.



the shape of the curve remains almost constant for the Control Group.

The distributions of gains made by the three experimental groups are graphically represented in Diagram 2. In these curves, O scores were presented alone, in ungrouped form, so as to show the exact number of children who made no improvement, as contrasted with those who gained or lost in the second test. In the Control Group, the portion of the curve falling at the minus end af the scale is distinctly larger than in the other two cases. The greatest number of O scores are likewise found in this group, while the least occur in the Praised Group. The mode of the minus scores for all three groups falls within the same range of scores, 0 to—5, while the mode of the plus scores occurs in the range of 0 to 5 Control Group, 5 to 10 of the Reproved Group, and spreads almost evenly from 0 to 15 in the case of the Praised Group.

When the results of this experiment are compared with those of similar studies, we find certain striking differences, mainly in the case of reproof. Gilchrist,[5] found that the Reproved Group made "no improvement in spite of practice. Their average score on the first test was 59.8 and on the second, 59.21. The Praised Group, on the other hand, improved by 79%. Their average score on the first test was 72.42 an the second, 129.5."

Gates and Rissland,[4] in their study based on the results of work in color-naming and motor co-ordination tests, found that in the co-ordination test, 89% of the Encouraged, 70% of the Discouraged, and 64% of the Repetition Groups gained. In the color-naming test, the percentages were smaller. Fifty-eight per cent of the Encouraged, 51% of the Discouraged, and 44% of the Repetition Groups did better on the second test.

Laird,[11] and[12] in his study based on the introspective reports of 92 college students, found that 26% of the high school students and 38% of the college students reported that they did better work following a public reprimand, while 76% of the high school and 84% of the college students improved their work following public praise.

The writer feels that these differences in the case of reproof can be explained on the basis of the differences in the conditions of the experiments. In the studies made by both Gilchrist and Gates, the situation was a laboratory one. The

subjects were taking the tests more or less as a favor to the experimenter, and they knew that the results of these tests would in no way affect their academic standing. The situation of the present study was entirely different. The children were in a typical school situation; they felt that the tests were an important matter, and that whatever scores they made would not only affect their school standing, but would be taken into account in determining promotions. Hence, reproof for failure was looked upon as a serious thing, and they reacted to it accordingly. In Laird's study, the situation is again somewhat different. One student was praised or reproved before the whole class, instead of a small group receiving the praise or reproof. This especially in the case of sensitive students, may to a great extent be responsible for the deterioration in work which followed reproof.

Summary

1: When the results for the 136 children within each of the experimental groups were studied as a whole, it was found that a greater percentage showed improvement following both praise and reproof than when the test was repeated without any preliminary comment. While 52% of the subjects showed some gain in the Control Group, 79% of the Praised and 80% of the Reproved Groups likewise improved their scores.

2. The actual gains made by the Praised and Reproved Groups, expressen in I. Q. units, were large enough and the P. E.'s of the difference small enough to say conclusively that a true difference did exist between these groups and the Control Group.

3. The difference in per cent gaining, and in the amount of gains made by the Praised and Reproved Groups was too small to be statistically reliable. What gain there was, however, seemed to be in favor of the Reproved Group.

2. *Grade Comparisons*

In order to study the influence of praise and reproof on children of different chronological ages, an attempt was made at first to subdivide the children of each group into age groups. This, however, proved to be impractical, since the groups thus formed contained too small a number of subjects to give reliable results. A second objection to the method was that in

the younger age groups, there was a too large percentage of superior children, and in the older age groups, a too large percentage of inferior children. To obviate these difficulties, the groups were then subdivided according to school grades, and the average of each grade was taken to represent the age of the children within the group. Thus, 14 years was used as representative of the 8th grade, 11.6 of the 5th, and 9.9 of the 3rd.

Table 7 shows the distribution of subjects according to grades.

TABLE 7.

DISTRIBUTION OF CHILDREN ACCORDING TO GRADES.
GRADE COMPARISONS.

	Control Group.	*Praised Group.*	*Reproved Group.*	*Combined.*
8th Grade.	47	47	47	141
5th Grade.	44	44	44	132
3rd Grade.	45	45	45	135
Totals.	136	136	136	408

The results of the first test for the three grades in each group are to be found in Table 8.

TABLE 8.

I.Q. SCORES. FIRST TESTS. GRADE COMPARISONS.

	Control Group.	*Praised Group.*	*Reproved Group.*
Averages.			
8th Grade.	102.49	102.59	102.85
5th Grade.	97.82	97.86	97.84
3rd Grade.	95.93	95.60	95.35
σ			
8th Grade.	20.07	21.60	19.66
5th Grade.	16.96	17.54	17.21
3rd Grade.	14.33	14.96	14.59
P.E. av.			
8th Grade.	1.96	2.11	1.92
5th Grade.	1.71	1.77	1.74
3rd Grade.	1.43	1.49	1.46

In the results presented above, it may be seen that the largest difference occurring between any of the groups is .58 I. Q. points (3rd Grade, Control-Reproved Group) and the smallest, .02 I. Q. points (5th Grade, Control-Reproved and Praised-Reproved Group). These differences, as a whole, are so small that it may be said with relative certainty that any difference between the groups is little more than a chance difference.

TABLE 9.

PERCENTAGE OF CHILDREN WHO GAINED, REMAINED THE SAME, OR LOST IN THE THREE GROUPS. GRADE COMPARISONS.

	Control Group.	*Praised Group.*	*Reproved Group.*
% Gaining.			
8th Grade.	46.8	91.5	80.8
5th Grade.	52.3	65.9	77.3
3rd Grade.	57.8	77.8	82.2
% Same.			
8th Grade.	8.5	0	4.3
5th Grade.	13.6	6.8	9.1
3rd Grade.	33.8	17.8	11.6
% Losing.			
8th Grade.	44.7	8.5	14.9
5th Grade.	34.1	27.3	13.6
3rd Grade.	8.9	4.4	6.2

In Table 9 are given the percentages of subjects in each grade who improved, remained the same, or lost as a result of practice, praise or reproof. Fewer of the 8th grade children are affected by practice than 5th, and fewer 5th and 3rd. This may be due to the fact that the older children, who have necessarily had more tests than the younger ones, could adapt themselves to the test situation better in the first test, and thus did not gain so much from merely repeating the test. On the other hand, it might well have been the result of boredom on the part of the older children who saw no use in taking a test so nearly like the one they took only the week before.

Using the Control Group as a standard for the percentage of children who could be expected to improve from practice, the results show that in the 8th grade, 45% of the children gained under the influence of praise, while only 14% in the 5th grade and 20% in the 3rd seemed to be affected. As a result of being reproved, 34% of the 8th grade children gained, 25% of the 5th, and 24% of the 3rd. Thus, more children in the two lower grades seemed to be influenced by reproof than by praise, while more children in the 8th grade were influenced by praise.

The average amount of gain made by each grade in the three different groups may be seen in Table 10. As in the case of the percentage of children gaining in each group, the average amount of gain, over and above practice, for the 8th grade is decidedly greater than for the other two. The net gain, due to praise, is 8.72 I. Q. points for the 8th grade, and only 2.91 and 3.11 points for the 5th and 3rd grades respectively. The net gain, due to reproof, for the 8th grade was 7.89 I. Q. points—slightly lower than for praise—while for

the 5th and 3rd grades it is 4.37 and 2.98 points. The 5th grade shows an average improvement of 1.46 points more when reproved than when praised, while the 3rd grade shows practically no difference, the average improvement when praised being .13 points greater than when reproved. A consideration of the range of gains shows that the greatest gain in the Control Group was made by an individual in the 3rd grade; in the Praised Group, by the 8th grade; and in the Reproved Group, also by the 8th. The greatest loss in the Control Group occurred in the 8th grade, in the Praised Group in the 5th grade, and in the Reproved Group in the 8th grade again.

TABLE 10.

RESULTS EXPRESSED IN TERMS OF INDIVIDUAL GAINS.
GRADE COMPARISONS.

	Control Group.	*Praised Group.*	*Reproved Group.*
Averages.			
8th Grade.	–.02	8.70	7.87
5th Grade.	1.45	4.36	5.82
3rd Grade.	2.24	5.35	5.22
σ			
8th Grade.	7.43	6.81	8.72
5th Grade.	5.48	8.05	6.19
3rd Grade.	5.79	5.66	5.38
P.E. av.			
8th Grade.	.73	.66	.85
5th Grade.	.55	.81	.62
3rd Grade.	.58	.57	.54
Range.			
8th Grade.	16 to –25	31 to –3	32 to –14
5th Grade.	14 to –11	30 to –9	20 to –6
3rd Grade.	18 to –12	17 to –6	15 to –10

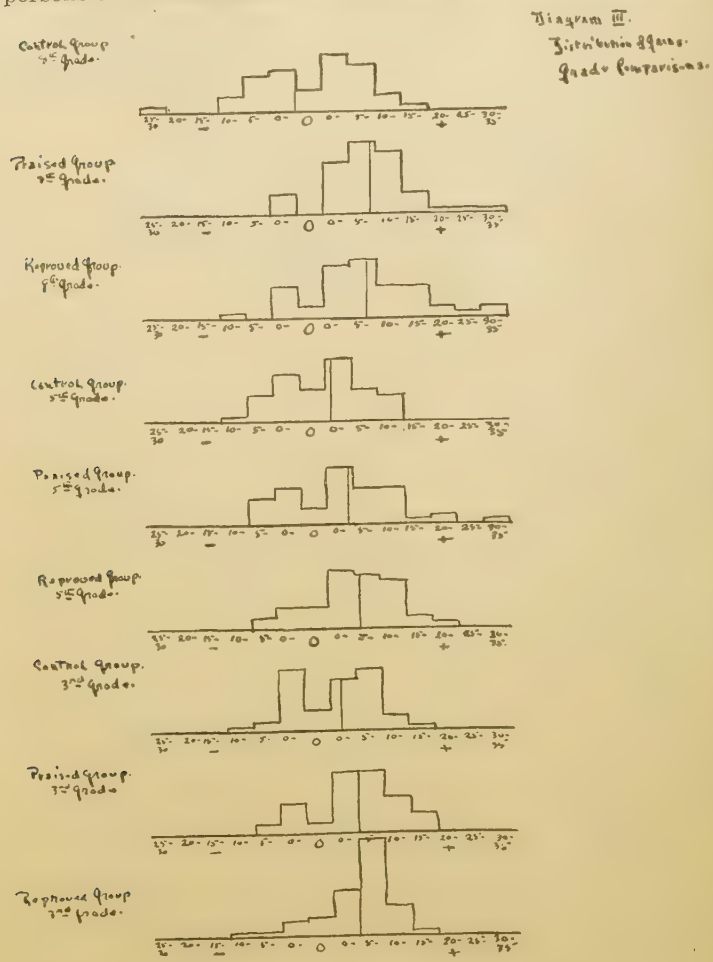
RELIABILITY OF DIFFERENCES.

Groups.	*Actual Difference.*	*P.E. Difference.*	*Difference in P.E. Units.*	*Chances that true diff. is above 0.*
8th Grade.				
Praised-Control.	8.72	.98	8.89	10,000 in 10,000.
Reproved-Control.	7.89	1.12	7.04	10,000 in 10,000.
Praised-Reproved.	.83	1.07	.77	6,982 in 10,000.
5th Grade.				
Praised-Control.	2.91	.97	3.00	9,785 in 10,000.
Reproved-Control.	4.37	.83	5.26	10,000 in 10,000.
Praised-Reproved.	–1.46	1.02	1.43	8,326 in 10,000.
3rd Grade.				
Praised-Control.	3.11	.81	3.83	9,951 in 10,000.
Reproved-Control.	2.98	.79	3.77	9,945 in 10,000.
Praised-Reproved.	.13	.78	.17	5,456 in 10,000.

Diagram 3 pictures graphically the gains for each grade in the three groups. Here, it is very evident that the 8th grade shows a decided shift in its scores towards the upper part of

the scale, and away from O. In the case of the two lower grades, on the other hand, the curves for the Praised and Reproved Groups are only slightly dissimilar to that of the Control Group.

Since other studies similar to this one have used as subjects, college students of approximately equal ages, no reference has been made to the effect of praise and reproof on persons of different ages. Nevertheless, the question has been



discussed theoretically, although no experimental evidence is given in support of the statements made. Albertine Richards[17], in discussing the question of motive in education, says that "since motivation depends on instinct and the desire for self-realization, it is obvious that motivation changes its character from individual to individual, and from age to age." Norsworthy and Whitley[18], when making suggestions for the best incentives to hold the child's attention, first of all stress the point that the teacher should "choose those natural to the child's stage of development."

Perhaps the best and most complete explanation of this sort has been given by McDougall[19]. Speaking of moral conduct and its development, he says: "We may roughly distinguish four levels of conduct, successive stages, each of which must be traversed by every individual before he can attain the next higher stage. These are (1) the stage of instinctive behavior modified only by the influence of the pains and pleasures that are incidentally experienced in the course of instinctive activities; (2) the stage in which the operation of the instinctive impulses is modified by the influence of rewards and punishments administered more or less systematically by the social environment; (3) the stage in which conduct is controlled in the main by the anticipation of social praise and blame; (4) the highest stage, in which conduct is regulated by an ideal of conduct that enables a man to act in the way that seems to him right regardless of the praise and blame of his immediate social environment."

If we are willing to accept this theory of development, as set forth by McDougall, we may then say that the children of the 8th grade, who were used as subjects for this study, were in the third stage of development, while the children of the 5th and 3rd grades were in the second stage, or in a transition stage from the second to the third. The difference in effectiveness of praise and reproof for children of different ages may thus in part be explained. Some simple reward in place of praise and some form of punishment in place of reproof would doubtless have been better forms of incentive to use for the children of the two lower grades.

Summary

When the results were considered from the point of view of the effect of age, as represented by the average age of the

children of the three grade groups, the following facts were brought out:

1. The 8th grade children were less affected by practice than were the 5th, and the 5th less than the 3rd.

2. The 8th grade children did better after being praised than after being reproved, while the opposite was true in the case of the 5th grade. The differences in the results for the 3rd grade groups were so small as to be little more than chance differences, but whatever slight difference did exist was in favor of reproof.

3. These results suggest that praise and reproof are not the best forms of incentive which might be used for younger children. For them, reward or punishment of some sort would perhaps be more effective and would at the same time appeal more strongly to the children themselves.

3. *Sex Comparisons*

The fact that girls of the elementary school ages are, as a whole, more serious about their school work than are boys of the same ages, cannot be denied. They not only study harder, but they take more pride in their successes, and are more distressed by failures. Hence, the natural inference is that they would be more sensitive to praise and reproof, when this praise and reproof are given in connection with their school work, and would accordingly be more influenced by them than would boys of similar ages. With this supposition in mind, the writer chose subjects of both sexes, with the hope that through an objective measurement of this sort, some solution to the question might be offered.

Table 11 shows the distribution of children in each grade, when the boys are separated from the girls.

TABLE 11.

DISTRIBUTION OF CHILDREN. SEX COMPARISONS.

	BOYS.			*GIRLS.*		
	Control Group.	*Praised Group.*	*Reproved Group.*	*Control Group.*	*Praised Group.*	*Reproved Group.*
8th Grade.	25	25	25	22	22	22
5th Grade.	24	24	24	20	20	20
3rd Grade.	21	21	21	24	24	24
Totals.	70	70	70	66	66	66

TABLE 12.

I.Q. Scores. First Tests. Sex Comparisons. (Total Results.)

	Control Group.		Praised Group.		Reproved Group.	
	Boys.	*Girls.*	*Boys.*	*Girls.*	*Boys.*	*Girls.*
Averages.	104.54	92.60	104.42	92.74	104.92	92.19
σ	13.61	17.98	14.36	19.38	14.15	18.88
P.E. av.	1.09	1.48	1.15	1.59	1.13	1.56

TABLE 13.

Percentage of Children Who Gained, Remained the Same, or Lost in the Three Tests.

Sex Comparisons. (Total Results.)

	Control Group		Praised Group.		Reproved Group.	
	Boys.	*Girls.*	*Boys.*	*Girls.*	*Boys.*	*Girls.*
% Gaining.	42.9	62.1	80.0	77.3	87.1	72.7
% Same.	11.4	9.1	5.7	1.5	1.5	12.2
% Losing.	45.7	28.8	14.3	21.2	11.4	15.1

TABLE 14.

Results Expressed in Terms of Individual Gains.
Sex Comparisons. (Total Results.)

	Control Group.		Praised Group.		Reproved Group.	
	Boys.	*Girls.*	*Boys.*	*Girls.*	*Boys.*	*Girls.*
Averages.	.35	2.26	7.25	5.05	8.14	4.58
σ	6.89	6.72	8.14	5.76	7.14	6.23
P.E. av.	.55	.55	.65	.48	.57	.51
Range.	16 to –13	18 to –25	31 to –7	19 to –9	30 to –10	32 to –14

Reliability of Differences.

Groups.	*Actual Difference.*	*P.E. Difference.*	*Difference in P.E. Units.*	*Chances that true diff. is above 0.*
Boys.				
Praised-Control.	6.90	.85	8.12	10,000 in 10,000.
Reproved-Control.	7.79	.79	9.86	10,000 in 10,000.
Praised-Reproved.	–.89	.86	1.03	7,563 in 10,000.
Girls.				
Praised-Control.	2.79	.73	3.82	9,950 in 10,000.
Reproved-Control.	2.32	.75	3.09	9,814 in 10,000.
Praised-Reproved.	.47	.70	.67	6,743 in 10,000.

In tables 12, 13 and 14 are shown the total results for boys and girls in each of the three experimental groups. Table 12 represents the results of the first testing, giving the initial I. Q. averages for each group: Table 13 shows the percentage of children who improved, remained the same, or lost, as a result of practice, praise or reproof: while Table 14 gives the amount of the gain, and the reliability of the gain for each group. Table 15 shows a direct comparison of the boys and girls within the three groups.

Certain considerations may be called to the reader's attention in connection with each of these tables. Table 12, show-

ing the I. Q. scores, presents two noteworthy facts. In the first place, the boys, although presumably chosen from grade sections similar to those of the girls, have, on the average, a higher I. Q.; the difference in this case being 12 points. Nevertheless, the average scores for both sexes fall within the group 90-110, the limits of "Average," as set by Terman[21]. In the second place, the differences between the three experimental groups for both boys and girls are small enough that the probability of a true difference of over 0 existing between them is very slight.

Several important results are shown in Table 13, in which are given the percentages of children who gained, remained the same, or lost. In the Control Group, 62% of the girls and only 43% of the boys gained. Using the results of the Control Group as a standard for practice effects, and subtracting that from the percentage who improved in the other two groups, we find that in the case of the boys, 37% more of the Praised, and 44% more of the Reproved Groups showed improvement than in the case of the Control Group. The girls, on the other hand, had a much smaller percentage improving in both groups. Fifteen per cent of the Praised and only 11% of the Reproved groups did better.

Here, it seems, we have a true sex difference. The girls, when no incentive is present, make more improvement than do the boys; while the boys, on the other hand, do decidedly better than the girls when praised or reproved. Although the differences between the percentages improving as a result of praise or reproof, are small, there is, nevertheless, a slight tendency for the boys to be more influenced by reproof, and the girls more by praise.

Table 14, showing the actual amount of gain made, again emphasizes the superiority of the girls in the Control Group and of the boys in the other two groups. In the case of the boys, the actual difference between the Control and Praised, and the Control and Reproved Groups is great enough to be statistically reliable. The differences for these groups for the girls are much smaller. Although not quite large enough to be absolutely reliable, statistically, they are, however, reliable enough to say with almost certainty that a true difference does exist, even if this be a small one. An interesting difference is again found between the Praised and Reproved Groups. The boys show a tendency to be more influenced in

the latter case, and the girls in the former. These differences are not great enough to be more than an interesting indication of a true difference.

TABLE 15.

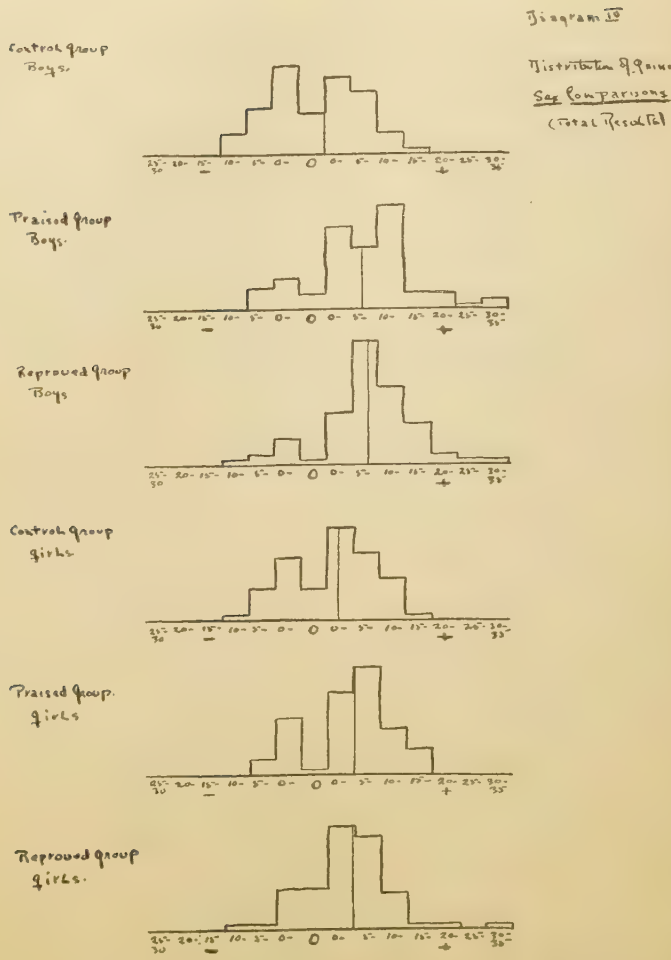
COMPARISONS WITHIN THE THREE EXPERIMENTAL GROUPS.
SEX COMPARISONS. (TOTAL RESULTS.)

Groups.	*Actual Difference.*	*P.E. Difference.*	*Difference in P.E. Units.*	*Chances that true diff. is above 0.*
Control.				
Boys-Girls.	–1.91	.78	2.45	9,508 in 10,000.
*Praised.**				
Boys-Girls.	4.11	.81	5.07	10,000 in 10,000.
*Reproved.**				
Boys-Girls.	5.47	.77	7.10	10,000 in 10,000.

*Note. Practice effect, as measured by the average improvement in the Control Group, was subtracted before the actual differences were determined.

In Table 15 are to be found comparisons of boys and girls within each group. The actual differences in average scores are given, together with the reliability of these differences. The difference in the average gains made by boys and girls in the Control Group was 1.91 points, in favor of the girls, with 95 out of 100 chances that the true difference was greater than 0. The boys, on the other hand, proved to be superior to the girls in actual average gains made in both the Praised and Reproved Groups. In the former case, the difference was 4.11 points, and in the later, 5.47, large enough in both cases to be statistically reliable.

In Diagram 4, the gains for the boys and girls in the three groups are shown graphically. The curves for the Control Groups of both sexes are practically identical. Both contain two peaks which occur in the same groups of scores. The one is in the group 0 to –5, and the other, 0 to 5. Such a close similarity in the form of the curves is not found for the Praised or Reproved Groups. In the former case, the curve of gains for the boys shows only a slight mode in the minus scores (in the group 0 to –5) while the curve for the girls has a decided mode in this part of the range of gains. The mode in the plus scores for the boys occurs in the group 10 to 15, and for the girls, 5 to 10. In the latter case, the point of greatest frequency in the scores for the boys is in the same position as in the Praised Group, while the girls' scores show no such mode. In the plus scores, the mode occurs in the group 5 to 10 for boys, and for the girls, it spreads about evenly from 0 to 10.



In order to determine the effect of the age factor on the performance of the different sexes, the groups were next subdivided according to grade distribution. Referring back to Table 3, in which are given the average ages for boys and girls in each grade, it appears that in every case, the average age for girls is slightly above that for boys. For example, in the 8th grade, the average age for boys was found to be 13.67 years, and for girls, 14.29 years: in the 5th grade, 11.12

years for boys, and 12.23 for girls: and in the 3rd, 9.12 for boys and 9.87 for girls. However, the overlapping in the groups is so great and the average difference so small that the two sexes within each grade may be considered equal in age.

In the 2 following tables, Tables 16 and 17, are presented the results when divided according to the grade distribution of the subjects.

TABLE 16.

PERCENTAGE OF CHILDREN WHO GAINED, REMAINED THE SAME, OR LOST IN THE THREE GROUPS.

	SEX COMPARISONS. (GRADE DIVISIONS.)					
	Control Group.		*Praised Group.*		*Reproved Group.*	
	Boys.	*Girls.*	*Boys.*	*Girls.*	*Boys.*	*Girls.*
% Gaining.						
8th Grade.	56.0	36.4	96.0	86.4	88.0	72.7
5th Grade.	37.5	70.0	75.0	55.0	91.7	60.0
3rd Grade.	33.3	79.2	66.7	87.5	80.9	83.4
% Same.						
8th Grade.	8.0	9.1	0	0	0	9.1
5th Grade.	16.7	10.0	8.3	5.0	0	20.0
3rd Grade.	9.5	8.3	9.5	0	4.8	8.3
% Losing.						
8th Grade.	36.0	54.5	4.0	13.6	12.0	18.2
5th Grade.	45.8	20.0	16.7	40.0	8.3	20.0
3rd Grade.	57.2	12.5	23.8	12.5	14.3	8.3

Table 16 shows the percentage of children in each grade in the three experimental groups who gained, remained the same, or lost. In the Control Group, more of the 8th grade boys improved than in the 5th or 3rd, while more of the 3rd and 5th grade girls improve than do those of the 8th grade. Taking into account the practice effect as determined by the Control Group as standard, the results show that when praise is used as an incentive, 40% of the 8th grade boys improve, 37% of the 5th, and 34% of the 3rd. The difference in percentages is so small that it seems justifiable to say that for the boys used in this study, praise was an equally effective incentive for all ages. In the case of the girls, on the contrary, there is a marked difference according to differences in age. The older girls, as represented by the 8th grade, have 50% more of the group improving under the influence of praise, when practise is accounted for, while 15% less of the 5th grade girls show improvement than in the case of the Control Group, and only 9% more in the 3rd grade show any improvement. Here again we have a striking sex difference. The boys, as a whole, regardless of age, do decidedly better when praised,

while only the older girls improve their performance, the other two grades showing negative or only slightly positive improvement.

Reproof seems to be a more effective incentive for 5th and 3rd grade boys, while less effective for those of the 8th grade than praise. Thirty-two per cent of the 8th grade boys, while 54% of the 5th and 48% of the 3rd improved their scores. Eighth grade girls, as was found to be true of boys of the same grade, do slightly worse when reproved than when praised. Here, 37% show improvement as contrasted with 50% in the Praised Group. No appreciable difference is seen in the performance of the 5th or 3rd grade girls in this group, as contrasted with the results of the Praised Group. Ten per cent of the former do worse and 4% of the latter do better after reproof. Thus, it seems as if the younger girls, as represented by the 5th and 3rd grade of this study, show little, if any, improvement following praise or reproof, other than a chance variation.

Turning now to Table 17 in which are presented the average amounts of gains for the groups, we find that in the Control Group, the 8th grade girls lose 1.50 points on re-testing while the boys of that grade gain 1.28 points. The boys in the 5th and 3rd grades show little or no practice effect, making gains of .38 and —1.24 points respectively. The girls, on the other hand, gain considerably in these grades. The average gain for the former is 2.75 points and for the latter, 5.29 points.

TABLE 17.

RESULTS EXPRESSED IN TERMS OF INDIVIDUAL GAINS.
SEX COMPARISONS. (GRADE DIVISIONS.)

	Control Group.		*Praised Group.*		*Reproved Group.*	
	Boys.	*Girls.*	*Boys.*	*Girls.*	*Boys.*	*Girls.*
Averages.						
8th Grade.	1.28	–1.50	10.32	6.82	10.44	4.95
5th Grade.	.38	2.75	6.75	1.50	8.29	2.85
3rd Grade.	–1.24	5.29	4.19	6.38	5.24	5.68
σ						
8th Grade.	7.78	6.66	8.04	5.58	8.01	8.61
5th Grade.	7.42	6.02	8.78	5.81	6.02	4.55
3rd Grade.	4.52	4.99	5.51	5.84	6.20	4.22
P.E. $_{av.}$						
8th Grade.	1.04	.95	1.08	.80	1.07	1.23
5th Grade.	1.01	.90	1.18	.87	.82	.68
3rd Grade.	.66	.68	.81	.79	.91	.58
Range.						
8th Grade.	16 to –13	11 to –25	31 to –3	19 to –3	30 to –2	32 to –14
5th Grade.	11 to –11	14 to –8	30 to –7	14 to –9	20 to –5	12 to –6
3rd Grade.	8 to –12	18 to –4	13 to –5	17 to –6	13 to –10	15 to –3

The groups that were praised and reproved show also some interesting sex differences as affected by age. When praised, the boys, on the average, gain, over and above practice, 9.04, 6.37 and 5.43 points for the 8th, 5th and 3rd grades respectively. When reproved, their gains for these grades are 9.16, 7.91 and 6.48 points respectively. There is thus a tendency for the older boys to be more influenced by both praise than reproof, and for all of them to be slightly more influenced by the latter. The girls, when praised, showed average gains of 8.32, –1.25 and 1.09 points, and when reproved, 6.45, .10 and .39 points for the 8th, 5th and 3rd grades respectively. The 8th grade girls, following praise, seemed to be influenced much in the same manner as the boys of that grade, but less influenced when reproved. The younger girls, on the other hand, show little or no gain as the result of either of the incentives. For the 5th grade, reproof seems to be a slightly more effective incentive than praise, while the opposite holds true in the case of the 3rd grade girls.

Laird's study [11, 12] was the only one of this kind in which the question of sex differences in reactions to the incentives of praise and reproof was considered. Laird found that in the case of reproof, high school men and women tended to be influenced in practically the same way. Of the former, 23.5% did better, and of the latter 28.5% improved their work. College men, on the other hand, showed very different results from the women. While only 16% of them reported as having done better, 60% of the women responded with better work.

In the case of high school students, there were found to be only slight sex differences in response to praise and again greater differences for college students. Of the high school men, 78.6% improved their work and 73.3% of the women. In college, 94% of the women and 75% of the men were spurred on to better work.

The only explanation which the writer can offer for the results found in this study is that the girls, as a whole, do better work when no incentive is present than do boys, and hence, when an incentive is introduced, they are not able to increase their performance, no matter how hard they may try. Boys, on the other hand, tend, as a whole, to work below their maximum ability, so that when there is an incentive for them to do better, they can increase their output. The pedagogical significance of this is obvious. To obtain the best work from

a group of boys, similar to this group, it would seem that some incentive were necessary. Hence, the conclusion would be: use praise whenever possible rather than reproof, and especially in the case of boys is some incentive necessary if they are to work up to their maximum.

Summary

A. When the groups as a whole were studied, the following results were found:

1. A larger percentage of girls improved in the Control Group as the result of practice than did the boys. Sixty-two per cent of the former showed some gains and 43% of the latter.

2. Taking into consideration the practice effect, determined by the Control Group as standard, 37% of the boys were spurred on to better work by praise and 44% by reproof. The girls, on the other hand, seemed to be much less influenced. Only 15% did better following praise, and 11% following reproof.

3. The actual gains for the groups showed again that the girls improved more as a result of practice than did the boys, but were less influenced by both praise and reproof. The gain in the Control Group was .35 I. Q. points, on the average, for the boys, and 2.26 for the girls. In the Praised Group, the boys gained an average of 6.90 points, over and above practice, and the girls, 2.79. In the Reproved Group, the gain for the boys was 7.79 and for the girls, 2.32 points.

B. The results for grade distributions of the sexes showed that:

1. The older boys did better than the younger ones as the result of practice, while the opposite was true in the case of the girls.

2. In the Praised Group, the boys were affected in almost the same manner throughout the three grades, while the 8th grade girls alone seemed to be spurred on to better work.

3. When reproved, again only a slight difference was seen in the results of the boys. The younger girls did better than when praised, and the older girls slightly worse.

4. *Levels of Intelligence*

Pintner[16], in discussing his new Mental-Educational Tests in 1921, made the following statement: "As a matter of fact,

we shall see what we have long suspected and what we have only recently consciously realized, that the greatest amount of waste exists among the brighter pupils of a class or among the better schools in a school system. It is, as a rule, the more intelligent pupils who are working below capacity, even when they are keeping well up to the average of the class. We have been pushing and cramming the duller children, while the brighter ones have been allowed to loaf. The bright child is the most retarded in our schools. The dull child is the most accelerated." Commenting on this statement of Pintner's, Toops and Symonds[22] in an article, "What Shall We Expect of the A. Q.?", ask the following question: "Is is not a remarkable coincidence that the 'below' normal in intelligence are for the most part above in motivation, while the 'above' normal in intelligence are for the most part below average in motivation?"

With Pintner's results and the question of Toops and Symonds in mind, the writer tried to determine whether praise and reproof, two of the most commonly used forms of motivation in school work proved to be more effective in causing increased performance on the part of the more intelligent than on the part of the less intelligent child.

The subjects in each of the three experimental groups were subdivided into three classes—the superior, the average, and the inferior—using as basis for this division, the classification given by Terman[23]. All of the subdivisions of the "superior" children of his classification—the "genius," "very superior intelligence" and "superior intelligence"—were classified simply as "superior" for this study, and in like manner, all of the subdivisions of the "inferior,"—"dullness," "border-line deficiency" and "definite feeble-mindedness"—were combined into the "inferior" group. The "average" group remained the same as in his classification. The reason for condensing the groups in this manner was that in the groups at the extreme ends of the scale, there were found to be too few cases to give reliable results.

The revised classification, as used in this study, was:

Above 110 I. Q.—Superior.
Between 90-110.—Average.
Below 90 I. Q. —Inferior.

Table 18 shows the distribution of subjects in each of the groups, when classified as "superior," "average" and "in-

ferior." There are approximately the same number of "superior" and "inferior" children in each of the groups, when the combined groups are considered. The boys used in this study tend to fall more into the "superior" and "average" groups, while the largest number of girls are in the "average" and "inferior" groups.

TABLE 18.

DISTRIBUTION OF CHILDREN. LEVELS OF INTELLIGENCE.

	Control Group.			*Praised Group.*			*Reproved Group.*		
	Boys.	*Girls.*	*Combined.*	*Boys.*	*Girls.*	*Combined.*	*Boys.*	*Girls.*	*Combined.*
Superior.	23	13	36	22	13	35	22	11	33
Average.	40	19	59	40	18	58	39	23	62
Inferior.	7	34	41	8	35	43	9	32	41
TOTALS.	70	66	136	70	66	136	70	66	136

TABLE 19.

I.Q. Scores. First Test. Levels of Intelligence.

	Control Group.			Praised Group.			Reproved Group.		
	Boys.	Girls.	Combined.	Boys.	Girls.	Combined.	Boys.	Girls.	Combined.
Averages.									
Superior.	119.67	122.46	120.68	120.59	122.53	121.31	121.00	123.54	121.85
Average.	100.50	97.57	99.56	100.35	99.55	100.10	101.02	98.22	100.00
Inferior.	82.28	78.44	79.09	80.37	78.17	78.59	82.55	77.09	78.29
σ									
Superior.	7.24	7.61	7.47	7.79	8.04	7.94	7.43	8.03	7.70
Average.	5.32	6.04	5.68	5.03	5.87	5.31	5.38	6.21	5.88
Inferior.	4.24	7.82	7.47	5.37	8.26	7.88	4.92	7.57	5.89
$P.E._{av.}$									
Superior.	1.01	1.41	.83	1.11	1.49	.90	1.06	1.62	.90
Average.	.56	.93	.50	.53	.93	.47	.58	.87	.50
Inferior.	1.07	.90	.78	1.27	.93	.80	1.10	.90	.62

In Table 19 are the results of the first testing. The "superior" girls tend to have a slightly higher average I. Q. score than do the boys of that group, while the reverse is true of the "average" and the "inferior" groups.

In Table 20 are given the percentages of children who improved, remained the same, or lost. In the Control Group, the "inferior" children, both boys and girls, are the ones who show a decided tendency to improve when merely repeating the test. About 1/3 of the "superior" children did better, ½ of the "average," and ¾ of the "inferior." When praised for their achievement, 35% more of the "superior" children gained than in the Control Group. Of this number, 47% were boys and only 16% girls. In the case of the "average" children, 32% of the boys, 25% of the girls, and 31% of the combined groups did better after praise. "Inferior" children also showed improvement, but not so much so as in the case of the former two. Thirty-one per cent of the boys, 10% of the girls,

and 13% of the two combined improved their scores. Reproof seemed to be a stronger incentive for "superior" and "average" children, but this was not true for the "inferior" ones. Of the "superior" children, 56% of the boys, 9% of the girls, or 40% of the boys and girls combined gained. In the group of "average" children, 39% of the boys, 27% of the girls, or 35% of the combined section gained. Only 8% of the "inferior" children showed any improvement as the result of reproof: 2% of these were girls and 32% boys.

TABLE 20.

PERCENTAGE OF CHILDREN WHO GAINED, REMAINED THE SAME, OR LOST IN THE THREE GROUPS.

LEVELS OF INTELLIGENCE.

	Control Group.			*Praised Group.*			*Reproved Group.*		
	Boys.	*Girls.*	*Combined.*	*Boys.*	*Girls.*	*Combined.*	*Boys.*	*Girls.*	*Combined.*
% Gaining									
Superior.	30.4	46.1	36.1	77.3	61.6	71.4	86.4	54.5	75.8
Average.	47.5	47.4	47.4	80.0	72.3	77.6	87.2	73.9	82.3
Inferior.	57.1	76.4	73.2	87.5	85.7	86.1	88.9	78.1	80.5
% Same.									
Superior.	17.4	15.4	16.7	9.1	0	5.7	0	9.1	3.0
Average.	5.0	15.8	8.5	5.0	5.5	5.2	2.6	17.4	8.1
Inferior.	28.6	2.9	7.3	0	0	0	0	9.4	7.3
% Losing.									
Superior.	52.2	38.5	47.2	13.6	38.4	22.9	13.6	36.4	21.2
Average.	47.5	36.8	44.1	15.0	22.2	17.2	10.2	8.7	9.6
Inferior.	14.8	20.7	19.6	12.5	14.3	13.9	11.1	12.5	12.2

The results given in this table show that less of the "superior" children do better work when unmotivated than do the "inferior," while a much greater number of the "superior" children do better when praised or reproved than do the "inferior." Besides this, there are striking sex differences in the findings. The girls, both "superior" and "inferior" do better than the boys when merely repeating the test, while many more of both of the groups, in the case of the boys, improved their work as the result of praise and reproof than did the girls of the same groups.

Table 21 contains the actual gains made on the test, together with the reliability of the differences of the gains. The "superior" children of the Control Group, both the boys and girls, did worse on the second test than on the first. The "average" children did slightly better, especially in the case of the boys, while the "inferior" children made a decided improvement. In the Praised Group, the "superior" boys averaged 9.17 I. Q. points gain over and above practice, the girls 4.05, and combined, 7.27. The "average" boys gained

4.86 points, the girls 3.54 and together, 4.46. In the case of the "inferior" children of this group, the boys made a most remarkable gain—10.16 points,—the girls only 1.78 and combined, 3.30. Since the average obtained for the boys was based on so few cases, and since the P. E. of the average is so large, the writer feels that little confidence can be placed in it. From the averages obtained in the Reproved Group, it appears that both "superior" and "inferior" boys and girls showed less improvement when reproved than when praised. The average gain for the "superior" boys was 8.82, for the girls, 2.72, and combined 6.77 points. The "inferior" boys gained 6.07, the girls .95 and together, 1.94 points. The opposite of this was found to be true for the "average" children. Both boys and girls did better, on the whole, when reproved than when praised, the boys gaining 7.28, the girls 4.04 and combined, 6.04 points.

TABLE 21.

Results Expressed in Terms of Individual Gains. Levels of Intelligence.

	Control Group.			Praised Group.			Reproved Group.		
	Boys.	Girls.	Combined.	Boys.	Girls.	Combined.	Boys.	Girls.	Combined.
Averages.									
Superior.	—1.87	—1.54	—1.75	7.30	2.51	5.52	6.95	1.18	5.02
Average.	1.31	.57	1.07	6.17	4.11	5.53	8.59	4.61	7.11
Inferior.	1.71	4.44	3.97	11.87	6.22	7.27	7.78	5.39	5.91
σ									
Superior.	5.26	9.98	7.36	8.46	5.63	7.88	7.21	3.36	6.79
Average.	6.36	4.31	5.86	7.73	6.32	7.35	7.43	4.76	6.84
Inferior.	3.27	5.78	5.58	7.68	5.93	6.66	5.71	6.67	6.58
P.E. $_{av.}$									
Superior.	.73	1.85	.82	1.21	1.04	.89	1.03	.68	.79
Average.	.67	.66	.51	.82	.99	.65	.79	.66	.58
Inferior.	.83	.67	.58	1.82	.67	.68	1.27	.79	.69
Range.									
Superior.	6 to –13	13 to –25	13 to –25	31 to –7	14 to –8	31 to –8	18 to –10	11 to –14	18 to –14
Average.	16 to –11	9 to –10	16 to –11	25 to –7	17 to –3	25 to –7	30 to –9	20 to –1	30 to –9
Inferior.	6 to –6	14 to –8	14 to –8	24 to –2	19 to –9	24 to –9	20 to –2	32 to –6	32 to –6

Reliability of Differences.

Groups.	Actual Difference.	P.E. Difference.	Difference in P.E. Units.	Chances that true diff. is above 0.
Praised-Control. (Sexes Combined.)				
"Superior"	7.27	1.21	6.01	10,000 in 10,000.
"Average"	4.46	.83	5.37	10,000 in 10,000.
"Inferior"	3.30	.89	3.71	9,938 in 10,000.
Reproved-Control. (Sexes Combined)				
"Superior"	6.77	1.14	5.94	10,000 in 10,000.
"Average"	6.04	.77	7.84	10,000 in 10,000.
"Inferior"	1.94	.90	2.15	9,265 in 10,000.
Praised-Reproved. (Sexes Combined)				
"Superior"	.50	1.19	.42	6,115 in 10,000.
"Average"	–1.58	.87	1.81	8,905 in 10,000.
"Inferior"	1.36	.96	1.42	8,309 in 10,000.

Considering the differences in actual gains made by the "superior," "average" and "inferior" children in the combined

groups, and the reliability of these differences, certain very interesting facts are brought out. (The results for boys and girls were not studied separately because the number of "superior" girls and "inferior" boys was too small to compare with the "superior" boys and the "inferior" girls.) When the reliability of the differences between the Control and Praised and Control and Reproved Groups for the "superior" and "average" children was determined, it was found that there was absolute certainty that the true difference was above 0; and for the "inferior" children, there was practical certainty that this was the case. The differences between the average results for the Praised and Reproved Groups were too small to be reliable, but showed that there was a strong tendency for the "average" child to be more influenced by reproof, the "inferior" more by praise, while in the case of the "superior" child, the difference, in favor of praise, was too small to be little more than a chance one.

The "superior" child was then compared with the "inferior" within each of the 3 experimental groups. (Table 22.) The results of the Control Group and of the Reproved Group showed that there was a true difference between the two types of children in the gains made. The "inferior" child made an average gain of 5.72 points above the "superior" in the Control Group, while the "superior" child gained 4.83 points more than the "inferior" in the Reproved Group. With practical certainty, it was also shown that the "superior" child's gain of 3.97 over that of the "inferior," due to praise, was a true difference.

TABLE 22.

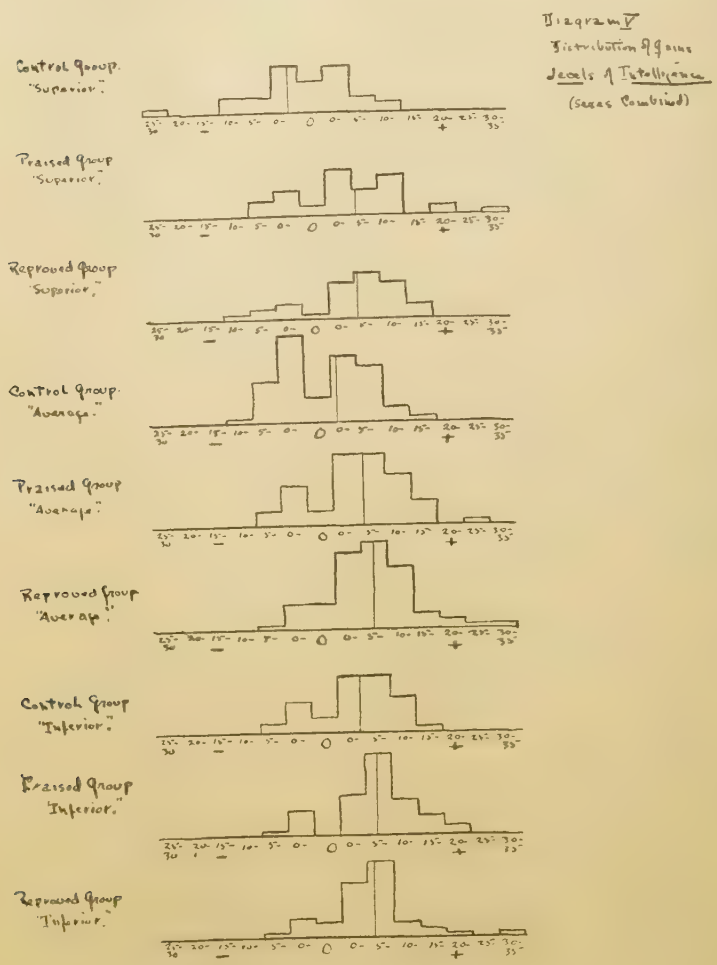
COMPARISONS WITHIN THE THREE EXPERIMENTAL GROUPS. LEVELS OF INTELLIGENCE. (SEXES COMBINED.)

Groups.	*Actual Difference.*	*P.E. Difference.*	*Difference in P.E. Units.*	*Chances that true diff. is above 0.*
Control.				
Superior-Inferior.	–5.72	1.00	5.72	10,000 in 10,000.
*Praised.**				
Superior-Inferior.	3.97	1.12	3.54	9,915 in 10,000.
*Reproved.**				
Superior-Inferior.	4.83	1.05	4.60	10,000 in 10,000.

*Note. Practice effect, as measured by the average improvement in the Control Group, was subtracted before the actual differences were obtained.

The curves of gains for each of the three subdivisions of the experimental groups are shown in Diagram 5. In the Control Group curves for the "superior," "average" and "inferior"

children, there are to be found slight modes in both the plus and minus scores. The peak of the minus scores comes in the group 0 to –5, and in the plus, 0 to 5 for the "superior" children, and for the other two classes, 0 to 10. In the curves for the Praised Groups, one sees less tendency for a mode to occur in the minus scores than in the previous cases. In all three curves, the point of greatest frequency falls almost



equally from 0 to 15. In the Reproved group, the curves for the "superior" and "average" children are almost identical in form with those of the Praised Groups. The curve for "inferior" children, on the other hand, shows the point of greatest frequency in the scores 0 to 10.

Again, a comparison was made between the results of this study and those of previous writers. Gilchrist[5] found that "the Reproved Group showed a noteworthy peculiarity. The members of the class who had done well in the first test did not do so well on the second. For example, the average score of the five highest in the first test was 14.0 and on the second, 8.13. Those who did not do so well in the first test made better scores in the second. The average score for the five lowest was 3.5 in the first test, and 7.25 in the second." His comment on this finding was: "The results in the second test show that the 'good' students are more influenced by the comment than the 'poor' students were. It may be that the former were unaccustomed to disparaging remarks, whereas the latter were inured to them." Gilchrist does not comment on the results found in the case of the Praised Group. If we study these, however, we find that there was improvement throughout. The five highest students made an average score on the first test of 13.9 points, and on the second, 19.1. The five lowest gained from an average score of 3.0 points to one of 12.4 points.

Gates and Rissland[4] also studied the effect of encouraging and discouraging comments upon groups differing in their initial abilities. In the co-ordination test, for example, the 8 highest had 63% of the number improving, the 10 middle ones, 100% and the same of the 8 lowest, when an encouraging comment preceded the second trial. When discouraged, 72% of the highest, 89% of the middle, and 57% of the lowest did better. In order to determine just how much of the improvement was due to practice, the results for the Repetition Groups were also given. It was found that 63% of the highest, 45% of the middle, and 87% of the lowest improved their work when no incentive was present. Their conclusion to this study is this: "Though the apparent difference may be a spurious one, it recalls a quotation from Terman, 'that in general, the poorer the response, the better satisfied one should appear to be with it.'"

Summary

1. Our results seem to accord with the findings of Pintner and of Toops and Symonds. The "inferior" children appear to be working approximately up to their maximum capacity, regardless of the presence of external motivation. This is proved by the fact that 73% of them improve in the Control Group, while only 36% of the "superior" and 47% of the "average" children show similar gains.

2. Taking the results of the Praised and Reproved Groups on their face value, it seems as if the "inferior" and "average" children rank above the "superior" both in the per cent gaining and in the actual amount of the average gains. But, since the gains due merely to practice must be taken into consideration before such conclusions may be drawn, we find that this is not true. It is the "superior" children who stand first, the "average" next, and the "inferior" last.

3. Sex comparisons show that the "superior" and "inferior" girls of the Control Group improve more from practice than do the boys of the same groups, while the "average" ones are about equal. Again, in the case of the "average" children of the Praised and Reproved Groups, the differences between boys and girls were slight. In the two other groups, on the contrary, the boys showed a decidedly greater tendency to improve their work as a result of external incentives than did the girls with whom their results were compared.

4. Comparisons of "superior" with "inferior" children within each group showed the differences in each case to be large enough to be statistically reliable ones. In the Control Group, the difference was in favor of the "inferior" children, who gained most by practice, but in the other two groups, the difference was in favor of the "superior."

5. The results of this study show that praise and reproof are incentives which may be used to get the "superior" child to exert greater effort and to bring his work up to its maximum point. The two incentives seem to be of equal value, with a slight superiority in favor of praise. The "average" children were less strongly influenced by both incentives, and the "inferior" children decidedly less so than either of the others. Reproof, in the case of the "average" children produced slightly better results, while the "inferior" children resembled the "superior" ones in that they did better following praise.

6. It does not seem fair to say that the "inferior" and

"average" children are, on the whole, less influenced to do better work as the result of praise and reproof than are the "superior" ones. These incentives probably mean as much to them as to "superior," but, since they are, as a rule, working up to their maximum capacity, they cannot increase their scores on such a test as this, no matter how hard they may try.

5. *Teachers' Ratings*

The tendency for teachers to overemphasize certain character traits and to underestimate such factors as age and environment in rating children for school work is now a recognized fact. A slow, timid child often is characterized as dull while a more responsive child of the same mental level is rated as average or even superior. It seemed to the writer that an interesting point in connection with the present study was the question of whether a child usually rated as "superior" in his school work was more inspired to do better work following praise or reproof than a child accustomed to being rated as "inferior." Knowledge of human behavior would lead us to suppose that the form of incentive most unlike the one he is accustomed to would act as a stimulus to call forth increased effort and, as a consequence, increased performance.

In order to determine which children were rated as "superior," "average" or "inferior" by their teachers on the basis of their school work, each teacher was given a list of her pupils with the following instruction:

"Please rate each of the children whose names are listed below, on the basis of the work they do in school and not on the basis of their conduct. Please use the following scale in making each rating:

"A"—Pupil whose school work is above average of the class, i.e., "superior."

"B"—Pupil whose school work is average for his class, i.e., "average."

"C"—Pupil whose school work is below the average of the class, i.e., "inferior."

Table 23 shows the number of children rated as "superior," "average" or "inferior" in each of the three groups. The boys, as a whole, tend to be rated more as "superior" and

"average," and the girls as "average" and "inferior." This was also found to be true when the I. Q.'s were used as the basis for classification.

TABLE 23.

DISTRIBUTION OF CHILDREN. TEACHERS' RATINGS FOR SCHOOL WORK.

	Control Group.			*Praised Group.*			*Reproved Group.*		
	Boys.	*Girls.*	*Combined.*	*Boys.*	*Girls.*	*Combined.*	*Boys.*	*Girls.*	*Combined.*
Superior.	33	30	63	39	31	70	37	27	64
Average.	24	18	42	23	17	40	26	22	48
Inferior.	13	18	31	8	18	26	7	17	24
TOTALS.	70	66	136	70	66	136	70	66	136

In Table 24 are to be found the percentages of children who gained, remained the same, or lost in each group. Fewer "superior" children, it may be seen, gained in the Control Group than "average," and fewer "average" than "inferior." Taking into account the percentage improving as the result of practice, 28% more of the "superior," 30% more of the "average," and 20% more of the "inferior" children gained as a result of praise. This is contradictory to what one might expect, and the only possible explanation, it seems to the writer, is that the "inferior" child is more often working up to his capacity, even when no incentive is present, and is therefore incapable of improving his score in the second trial, regardless of his desire to do so.

Certain interesting sex differences are brought out in these results. The boys, when classed as "superior," "average" or "inferior" respond in almost the same way to praise. Thirty-five per cent of the "superior," 37% of the "average" and 34% of the "inferior" gained. In the case of the girls, on the other hand, praise seemed to be less effective for the "inferior" than for the "average" or "superior." Only 11% of the "inferior" gained while 21% of the "average" and 20% of the "superior" did better.

Reproof for the boys was more influential in bringing about increased work when used in the case of the "superior" boys than in the case of the "average" or "inferior." Fifty-three per cent of the "superior" ones improved and only 35% of the "average" and 32% of the "inferior." Except for the "superior" boys, where reproof seemed to be the more effective incentive, there was almost no difference in the value of the two. The same was not found to be true in the case of the girls.

TABLE 24.

PERCENTAGE OF CHILDREN WHO GAINED, REMAINED THE SAME, OR LOST IN THE THREE GROUPS. TEACHERS' RATINGS FOR SCHOOL WORK.

	Control Group.			*Praised Group.*			*Reproved Group.*		
	Boys.	*Girls.*	*Combined.*	*Boys.*	*Girls.*	*Combined.*	*Boys.*	*Girls.*	*Combined.*
% Gaining.									
Superior.	39.4	56.7	47.6	74.4	77.4	75.7	91.9	81.5	87.5
Average.	45.8	55.6	50.0	82.6	76.5	80.0	80.8	72.8	77.1
Inferior.	53.8	72.2	64.5	87.5	83.3	84.6	85.7	58.8	66.7
% Same.									
Superior.	12.1	13.3	12.7	7.7	0	4.3	0	7.4	3.1
Average.	0	11.1	4.8	4.3	5.9	5.0	3.9	13.6	8.3
Inferior.	23.1	5.6	12.9	0	0	0	0	17.7	12.5
% Losing.									
Superior.	48.5	30.0	39.7	17.9	22.6	20.0	8.1	11.1	9.4
Average.	54.2	33.3	45.2	13.1	17.6	15.0	15.3	13.6	14.6
Inferior.	23.1	22.2	22.6	12.5	16.7	15.4	14.3	23.5	20.8

While 25% of the "superior" and 17% of the "average" did better work after being reproved, 13% of the "inferior" did poorer work. Contrasting praise and reproof, we find that reproof is more effective for the "superior" girls as was the case when the results of the "superior" boys were studied. Reproof, on the other hand, was much less effective for the "inferior" girls than was praise, while the "average" group responded in much the same way to both.

When both sexes were combined, 40% of the "superior" children, 27% of the "average" and only 2% of the "inferior" showed gains over the percentage improving in the Control Group. Here we have a true difference and an anticipated one. The "inferior" children in the Praised Group showed an almost equal percentage gaining with those of the other two groups. In the case of reproof, however, the incentive appealed less forcibly to the "average" than to the "superior," and almost none at all to the "inferior."

It seems, then, from these results, that children who usually do poor work in school can be stimulated to improve by praise and not by reproof, average children respond to both incentives in about the same way, while reproof is a more forceful incentive to use in the case of "superior" children.

The averages of the actual gains made, together with the differences between the groups and the reliability of these differences are shown in Table 25. Again, the Control Group shows that it is the "inferior" children, both boys and girls, who profit most by practice, while the "superior" children profit least. Praise seemed to spur on the "average" child

TABLE 25.

Results Expressed in Terms of Individual Gains. Teachers' Ratings for School Work.

	Control Group.			Praised Group.			Reproved Group.		
	Boys.	Girls.	Combined.	Boys.	Girls.	Combined.	Boys.	Girls.	Combined.
Averages.									
Superior.	—.33	1.07	.33	6.71	5.64	6.24	9.16	4.59	7.23
Average.	.54	.55	.54	7.91	5.94	7.07	6.77	4.41	5.69
Inferior.	1.15	3.94	2.77	8.37	5.00	6.04	8.14	3.70	5.00
σ									
Superior.	5.07	6.04	5.59	7.89	6.34	7.25	6.83	6.18	6.91
Average.	5.89	8.38	7.10	8.44	5.17	7.28	7.78	3.98	6.46
Inferior.	6.56	5.08	5.94	6.73	6.24	6.53	5.56	8.15	7.78
P.E. av.									
Superior.	.59	.74	.47	.85	.76	.58	.75	.80	.58
Average.	.81	1.32	.73	1.18	.84	.77	1.02	.57	.62
Inferior.	1.22	.80	.71	1.59	.99	.86	1.41	1.33	1.06
Range.									
Superior.	11 to –13	13 to –11	13 to –13	31 to –7	19 to –6	31 to –7	30 to –5	20 to –14	30 to –14
Average.	12 to –9	18 to –25	18 to –25	31 to –7	14 to –3	31 to –7	20 to –10	12 to –2	20 to –10
Inferior.	16 to –11	14 to –5	16 to –11	16 to –5	18 to –9	18 to –9	15 to –2	32 to –6	32 to –6

Reliability of Differences.

Groups.	Actual Difference.	P.E. Difference.	Difference in P.E. Units.	Chances that true diff. is above 0.
Praised-Control. (Sexes Combined)				
"Superior."	5.91	.74	7.98	10,000 in 10,000.
"Average."	6.53	1.06	6.16	10,000 in 10,000.
"Inferior."	3.27	1.11	2.95	9,767 in 10,000.
Reproved-Control. (Sexes Combined)				
"Superior."	6.90	.74	9.32	10,000 in 10,000.
"Average."	5.15	.95	5.42	10,000 in 10,000.
"Inferior."	2.23	1.28	1.74	8,798 in 10,000.
Praised-Reproved. (Sexes Combined)				
"Superior."	—.99	.82	1.21	7,927 in 10,000.
"Average."	1.38	.99	1.39	8,256 in 10,000.
"Inferior."	1.04	1.36	.76	6,959 in 10,000.

slightly more than the "superior," and decidedly more than the "inferior." The average gain, allowing for practice, was 6.53 I. Q. points for the "average" children, 5.91 for the "superior" and 3.27 for the "inferior." Reproof had its greatest effect on the "superior" children, less on the "average" and very much less on the "inferior." The average gain for the "superior" children, discounted for practice, was 6.90, for the "average," 5.15; and for the "inferior," 2.23 I. Q. points.

Some striking sex differences are to be found in these results, but considering the numbers on which the averages are based in so many cases, one can readily see that too much reliance cannot be placed on them. However, it is interesting to note in passing that in the Control Group, the "superior" boys made a negative gain, and the "superior" girls only one point positive one. The "average" boys and girls gained the same amount, while the "inferior" girls gained over three times as much as the boys of that class. Praise proved to be more effective for all three classes of boys than for corresponding

classes of girls. The same was true for reproof in all three of the classes.

Turning now to a direct comparison of the groups, when boys and girls are combined, we find the results summarized at the bottom of Table 25. The differences between the Control and Praised, and the Control and Reproved Groups for the children rated "superior" and "average" were great enough to be statistically reliable results. In the case of the "inferior" children, the differences were not large enough to be reliable ones. The gain of the Praised over the Control Group was large enough to say with practical certainty that there was a true difference between them, greater than 0, while the true difference between the Reproved and Control Groups had only 88 out of 100 chances of being greater than 0.

TABLE 26.

COMPARISONS WITHIN THE THREE EXPERIMENTAL GROUPS.
TEACHERS' RATINGS FOR SCHOOL WORK. (SEXES COMBINED).

Groups.	*Actual Difference.*	*P.E. Difference.*	*Difference in P.E. Units.*	*Chances that true diff. is above 0.*
Control.				
Superior-Inferior.	–2.44	.85	2.87	9,735 in 10,000.
*Praised.**				
Superior-Inferior.	2.64	1.03	2.56	9,579 in 10,000.
*Reproved.**				
Superior-Inferior.	4.67	1.20	3.89	9,956 in 10,000.

*Note. Practice effect, as measured by the average improvement in the Control Group, was subtracted before the actual differences were obtained.

In Table 26 are given the direct comparisons of the "superior" with the "inferior" children within the three experimental groups. There is found to be practical certainty, although not absolute certainty, in every case that the true difference is above 0. The gain in the Control Group is in favor of the "inferior" children, but in the other two groups, in favor of the "superior."

SUMMARY

Summing up the results of this section, we find that:

1. Children rated as "inferior" by their teachers gained more as the result of practice than did "average" or "superior" children.

2. The "average" children showed a slight tendency to do better work than the "superior," after being praised, and a decided tendency to do better than the "inferior" children.

3. Reproof had its greatest effect on the "superior" children, less on the "average," and distinctly less on the "inferior."

4. Sex comparisons showed that the "superior" and "inferior" girls gained more from practice than did similar groups of boys, while the "average" groups showed no difference. Both praise and reproof brought about better results in the case of the boys, regardless of their classification, than it did for any of these groups of girls.

5. The differences between the Control and Praised, and Control and Reproved Groups for "superior" and "average" children were large enough to be absolutely reliable, statistically, and practically reliable in the case of the "inferior" children of the Praised and Control Groups, and in the case of the "superior" and "inferior" children who were compared in each of the three experimental groups. The difference between the "inferior" children of the Control and Reproved groups, and between the Praised and Reproved Groups for children of all three sections, was large enough to indicate the direction in which the gain was to be found, but not large enough for statistical certainty.

6. *Race Comparisons*

It is generally agreed by students of race psychology that the Negro is especially susceptible to flattery or praise, and that by appealing to his vanity, he will exert every possible effort to show off to his best advantage. In schools where negro children are taught in the same classes with white children, the negro child rarely receives what slight amount of praise is given for good school work. This is not due so much to the fact that his work is inferior to that of the average white child, but may be traced to the presence of racial prejudice. The white child, on the other hand, seems to take greater pride in his school work than does the negro, and for that reason one is led to suppose that reproof would be the more effective form of incentive for him.

As the writer had an opportunity to secure subjects in schools where about 35% of the children are negroes, it seemed that an interesting phase of the problem would be to compare white and negro children of the same social status in order to determine whether these assumed racial differences

really exist, when subjected to some form of objective measurement.

The distribution of white and negro children in the three experimental groups is shown in Table 27. No attempt was made to separate the results according to school grades, because there were too few negro children in the 8th grade.

TABLE 27.

DISTRIBUTION OF CHILDREN. RACE COMPARISONS.

	Control Group.			*Praised Group.*			*Reproved Group.*		
	Boys.	*Girls.*	*Combined.*	*Boys.*	*Girls.*	*Combined.*	*Boys.*	*Girls.*	*Combined.*
Whites.	45	41	86	45	41	86	44	41	85
Negroes.	25	25	50	25	25	50	26	25	51
TOTALS.	70	66	136	70	66	136	70	66	136

TABLE 28.

I.Q. SCORES. FIRST TEST. RACE COMPARISONS.

	Control Group.			*Praised Group.*			*Reproved Group.*		
	Boys.	*Girls.*	*Combined.*	*Boys.*	*Girls.*	*Combined.*	*Boys.*	*Girls.*	*Combined.*
Averages.									
Whites.	105.53	97.49	101.69	104.91	97.32	101.29	106.98	97.39	102.35
Negroes.	102.76	84.64	93.70	103.56	85.24	94.40	101.46	83.68	92.74
σ									
Whites.	13.43	19.48	17.06	15.88	19.97	18.36	13.68	18.78	17.12
Negroes.	13.39	14.05	16.49	11.94	15.44	16.55	13.76	15.71	17.20
P.E. av.									
Whites.	1.34	2.04	1.23	1.59	2.09	1.33	1.38	1.97	1.24
Negroes.	1.79	1.88	1.56	1.60	2.07	1.57	1.81	2.11	1.61

In Table 28 are given the results for the first test in I. Q. scores. The negro boys average almost as high as the white boys in each group, while the negro girls are a little over 10 points lower than the white girls. When the combined results are considered, it may be seen that the white children have, on the average, an I. Q. score of about 8 points higher than the negroes. However, the combined scores of both races fall within the range considered as "average" by Terman[21], 90-110, so the objection cannot be raised that what differences may occur are due to the fact that the negro children belong to the "inferior" group while the white belong to the "average" or "superior" groups.

Table 29 shows the percentages of children who improved, remained the same or lost in each of the three groups. In the Control Group, the white children show a slight tendency to profit more by practice than do the negro. Praise, however, seems to appeal more strongly to the negro children. Forty

per cent of them improved their scores as a result of praise, while only 18% of the white children did better than in the Control Group. On the other hand, the white and negro children seemed to respond in a similar manner to reproof. Twenty-seven per cent of the white children showed improvement and 28% of the negroes. Taking the results as a whole, we may then say that the negroes do better in response to praise and the white children in response to reproof.

TABLE 29.

PERCENTAGE OF PERSONS WHO GAINED, REMAINED THE SAME, OR LOST IN THE THREE GROUPS. RACE COMPARISONS.

	Control Group.			*Praised Group.*			*Reproved Group.*		
	Boys.	*Girls.*	*Combined.*	*Boys.*	*Girls.*	*Combined.*	*Boys.*	*Girls.*	*Combined.*
% Gaining.									
Whites.	55.5	53.6	54.7	75.5	70.7	73.3	88.7	75.6	82.3
Negroes.	20.0	76.0	48.0	88.0	88.0	88.0	84.6	68.0	76.4
% Same.									
Whites.	2.2	9.8	5.8	8.9	2.5	5.8	0	4.9	2.3
Negroes.	28.0	8.0	18.0	0	0	0	3.9	24.0	13.8
% Losing.									
Whites.	42.3	36.6	39.5	15.6	26.8	20.9	11.3	19.5	15.4
Negroes.	52.0	16.0	34.0	12.0	12.0	12.0	11.5	8.0	9.8

The boys and girls of this study seemed to react in a different manner in each of the three groups. The white boys and girls had an almost equal percentage improving in the Control Group, while about 4 times as many negro girls did better than did the boys. Again, there was almost no sex distinction between the white children when praise was used, but the negro boys seem to have been much more influenced than the girls. Sixty-eight per cent of the boys gained and only 12% of the girls. The only sex difference between the white children was found in the case of reproof. Thirty-three per cent of the boys and 22% of the girls showed improvement over the number gaining in the Control Group. The negro boys showed little difference in their response to praise and reproof,—68% gained in the former case and 65% in the latter. The girls, on the contrary, did worse than the Control Group. While 68% of them showed some gain, this was 8% less than the percentage gaining in the Control Group.

In Table 30 are presented the averages of the actual gains made by the white and negro children. Little difference is seen in the results of the Control Group for the two races; the white children averaging .31 I. Q. points higher than the

negro. The negroes, however, show a gain of 5.82 points as the result of praise, while the white children fall slightly behind them with an average of 4.41. Reproof, on the other hand, brought about the reverse order. The white children made an average gain of 5.34 points, and the negro, 4.84. Comparing the results for boys and girls of both races, we find that the girls in the Control Group in both cases do better than do the boys, but the difference in the case of the white children is not so marked as in the case of the negro. The effect of praise was found to be greater for both white and negro boys than for the girls of the same races. The average gain made by the white boys, over and above practice, was 5.47 points, and for the negro boys, 9.72 points. The white girls gained 3.24 points and the negro girls only 1.92. Here, we have a decided sex difference, and as in the case of the Control Group, the difference between the negro children was greater than between the white children. A similar difference between the sexes was found when reproof was used as an incentive. The boys made almost equal gains, the average for the white being 7.76 points, and for the negro, 8.47. The average gains of the girls for both races were very much smaller. The white girls averaged 2.76 points above the gain made by the Control Group, and the negro, 1.16. Here again, the sex difference is greater for the negroes, but very much smaller than in the two former cases.

TABLE 30.

Results Expressed in Terms of Individual Gains. Race Comparisons.

	Control Group.			Praised Group.			Reproved Group.		
	Boys.	Girls.	Combined.	Boys.	Girls.	Combined.	Boys.	Girls.	Combined.
Averages.									
Whites.	.95	1.66	1.29	6.42	4.90	5.70	8.71	4.42	6.63
Negroes.	—1.28	3.24	.98	8.44	5.16	6.80	7.19	4.40	5.82
σ									
Whites.	6.18	7.42	6.81	7.83	6.67	7.33	7.44	6.74	7.39
Negroes.	4.98	4.87	5.46	7.38	5.41	6.67	6.33	4.47	5.68
P.E. av.									
Whites.	.62	.78	.49	.78	.70	.53	.75	.71	.54
Negroes.	.67	.65	.52	.99	.72	.63	.83	.60	.53
Range.									
Whites.	16 to -13	18 to -25	18 to -25	30 to -7	19 to -8	30 to -8	30 to -10	32 to -14	32 to.-14
Negroes.	12 to -12	14 to -8	14 to -12	31 to -6	17 to -9	31 to -9	26 to -3	15 to -1	26 to -3

Reliability of Differences.

Groups.	Actual Differences.	P.E. Difference.	Difference in P.E. Units.	Chances that true diff. is above 0.
White Children.				
Praised-Control.	4.41	.72	6.12	10,000 in 10,000.
Reproved-Control.	5.34	.73	7.32	10,000 in 10,000.
Praised-Reproved.	—.93	.76	1.22	7,946 in 10,000.
Negro Children.				
Praised-Control.	5.82	.82	7.09	10,000 in 10,000.
Reproved-Control.	4.84	.74	6.54	10,000 in 10,000.
Praised-Reproved.	.98	.82	1.19	7,890 in 10,000.

Comparing the reliability of the differences between the groups, we find that when the whites are taken alone or the negroes alone, the actual differences between the Control and Praised, and the Control and Reproved Groups are large enough to be statistically reliable. Between the Praised and Reproved Groups, the differences that are found to exist for both races are so small as to be little more than chance ones.

In Table 31 are seen the results of a direct comparison of the white and negro children within each of the three groups. The comparisons are made for boys and girls separately and in combination. When the combined results are considered, it is found that the differences are too small in each case to be statistically reliable; in fact, in two out of the three cases, the difference is little more than a chance difference. Comparing the negro boys with the white boys, and the negro girls with the white girls within each group, we again find that the differences are too small to have statistical reliability except in the case of the boys in the Control and in the Praised Groups. Here, the white boys do better as the result of practice and the

TABLE 31.

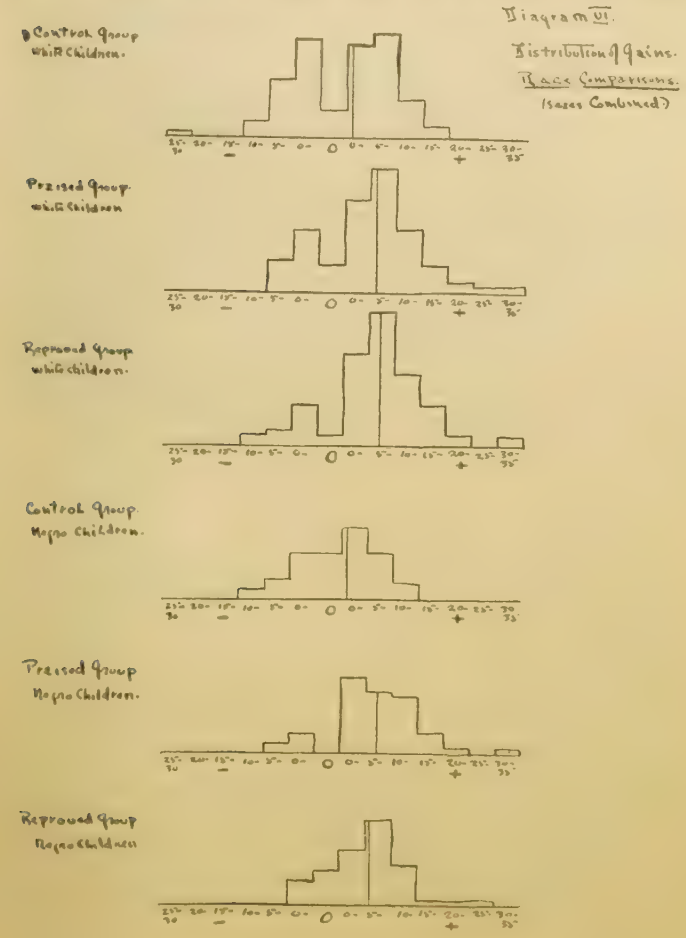
COMPARISONS WITHIN THE THREE EXPERIMENTAL GROUPS.
RACE COMPARISONS.

Groups.	*Actual Difference.*	*P.E. Difference.*	*Difference in P.E. Units.*	*Chances that true diff. is above 0.*
SEXES COMBINED.				
Control.				
Whites-Negroes.	.31	.72	.43	6,141 in 10,000.
*Praised.**				
Whites-Negroes.	–1.41	.82	1.72	8,770 in 10,000.
*Reproved.**				
Whites-Negroes.	.50	.75	.76	6,959 in 10,000.
BOYS.				
Control.				
Whites-Negroes.	2.23	.91	2.45	9,508 in 10,000.
*Praised.**				
Whites-Negroes.	–4.25	1.26	3.37	9,885 in 10,000.
*Reproved.**				
Whites-Negroes.	–1.71	1.12	1.53	8,489 in 10,000.
GIRLS.				
Control.				
Whites-Negroes.	–1.58	1.01	1.56	8,536 in 10,000.
*Praised.**				
Whites-Negroes.	1.32	1.00	1.32	8,133 in 10,000.
*Reproved.**				
Whites-Negroes.	1.60	.93	1.72	8,770 in 10,000.

*Note. Practice effect, as measured by the average improvement in the Control Groups, was subtracted before the actual differences were obtained.

negro boys as the result of praise. In the other comparisons, there is a tendency for the negro boys to be more influenced by reproof, for the negro girls to profit more from practice than the white, but for the white girls to be superior in their responses to both praise and reproof.

Diagram 6 shows the curves of gains for negro and white children in the three experimental groups. The results for the two sexes combined are given. The curves for the white



children show modes in both the plus and minus scores. The mode in the minus scores is greatest in the Control Group and smallest in the Reproved Group. In all three, however, it occurs in the group of scores 0 to –5. The mode in the plus scores likewise occurs in the same position in all three curves, 5 to 10. However, it is distributed fairly evenly from 0 to 10 in the Control Group, and from 0 to 15 in the Praised and Reproved Groups.

The curves of gains made by the negro children differ from those of the white children. In them, there appears to be only one mode, and that is found in the plus scores. In the Control and Praised Groups, this mode occurs in the scores 0 to 5, while in the Reproved Group, it is in the group 5 to 10. In the case of the minus scores, the distributions are fairly symmetrical, with no outstanding point of greatest frequency.

The results of this study are somewhat similar to the study made by McFadden and Dashiell[14] using the June Downey Will-Temperment test to determine race differences in tem permental qualities. They, too, found that there were almost no racial differences, and that what differences did exist were too small to be significant. We have found the same to be true in the responses to praise and reproof, as made by the two races. There is a slight tendency for praise to be a more effective incentive to be used for colored children, reproof for white children.

Summary

1. The white children, as a whole, have a higher average I. Q. score than do the negroes, but the averages for both fall within the range of "average." The boys of both races are about equal, while the score for negro girls is below that for white girls.

2. The percentages of children improving in each group show the white children to be superior to the negro in the Control Group, the negroes in the Praised Group, and the two races equal in the Reproved Group. There are sex differences within each of the groups. The white children are about equal in the percentages showing improvement in the Control and Praised Groups, while the boys are superior to the girls in the Reproved Group. The negro girls had a larger percentage of the group improving as the result of practice, but the boys

a much larger number as the result of both praise and reproof.

3. A study of the actual gains made shows certain differences which are not brought out when only the percentages improving are taken into consideration. When the combined results are compared, it may be seen that the white children are more affected by practice and reproof, and the negroes by praise. The sex comparisons of the two races show the girls of both races as doing better following practice than the boys, but the boys as being more spurred on to better work following both praise and reproof.

4. A direct comparison of the races within the three experimental groups showed the differences in average gains to be too small to be considered statistically reliable.

7. *Questionnaires*

In any study, in which the results are based on the introspections of the subjects, certain doubt is always felt about their reliability. This is especially true when the subjects happen to be children. One way to lessen the unreliability of such results is to make the introspections as objective as possible by asking very simple questions, and by allowing the subject to check one of the accompanying answers which is then taken as an indication of his reaction to the matter. Although we cannot ever be sure that a child is answering the question as he really thinks or feels, rather than as he thinks he ought to, nevertheless, we have perhaps as reliable material as could be obtained from any study of this kind. The writer, therefore, with this doubt in mind, presents the results of the questionnaire study. The answers given may be taken as an indication of the children's reactions to the situations in which they were placed, but, especially in the case of the younger children, too much importance is not to be attached to them.

Below are presented the three forms of questionnaires used. There was a different form for each of the three experimental groups, but within a group, the same form was used for children of different ages and school grades. Each child was given a questionnaire at the end of the second testing, and was asked to check his answers only after the question had been read aloud by the examiner.

Questionnaire for Control Group

1. When you finished the test last week, did you think that you had made a high score on it? YES. NO.
2. Do you think that you did better in this test than in the last one? YES. NO.
3. Did you try harder in this test than in the last one? YES. NO.
4. Did you enjoy taking this test more than the last one? YES. NO.

Questionnaire for Praised Group

1. When you finished the test last week, did you think that you had made a high score on it? YES. NO.
2. Were you surprised today when I told you that you had been selected from the whole group because of your very good work in the last test? YES. NO.
3. Do you think that you did better in this test than in the last one? YES. NO.
4. Were you encouraged to try to do better this time after I told you how well you did last time? YES. NO.
5. How long did this encouragement last? First Half of Test. Last Half of Test. Throughout Test.
6. Did you enjoy taking this test more than the last one? YES. NO.

Questionnaire for Reproved Group

1. When you finished the test last week, did you think that you had made a high score on it? YES. NO.
2. Were you surprised today when I told you that you had been selected from the whole group because of your very poor work in the last test? YES. NO.
3. Do you think that you did better in this test than in the last one? YES. NO.
4. After I told you how badly you had done in the first test, did this cause you to work harder and try to do better in this test? YES. NO.
5. How long were you discouraged by what I said to you? First Half of Test. Last Half of Test. Throughout Test.
6. Did you enjoy taking this test more than the last one? YES. NO.

From these, it may be seen that the questions are simple, direct and easily intelligible to the children. However, to insure understanding on the part of all of the children, the examiner read out each question to the group, and then told them to underline "yes" or "no," or whatever alternatives were used, so as to show their answer to that question.

TABLE 32.

PERCENTAGE OF ANSWERS GIVEN TO EACH QUESTION. QUESTIONNAIRE STUDY.

CONTROL GROUP.

Questions.	8th Grade.			5th Grade.			3rd Grade.			Totals.
	Boys.	Girls.	Combined.	Boys.	Girls.	Combined.	Boys.	Girls.	Combined.	
	%	%	%	%	%	%	%	%	%	%
1.										
Yes.	80.0	68.2	74.5	45.9	35.0	40.9	61.9	62.5	62.2	59.5
No.	20.0	31.8	25.5	54.1	65.0	59.1	38.1	37.5	37.8	40.5
2.										
Yes.	84.0	86.3	85.1	75.0	85.0	79.5	47.6	91.6	71.1	78.7
No.	16.0	13.7	14.9	25.0	15.0	20.5	52.4	8.4	28.9	21.3
3.										
Yes.	52.0	36.3	44.6	83.3	75.0	79.5	52.4	45.8	48.9	57.3
No.	48.0	63.7	55.4	16.7	25.0	20.5	47.6	54.2	51.1	42.7
4.										
Yes.	84.0	45.4	65.9	79.1	60.0	70.4	66.7	62.5	64.4	66.9
No.	16.0	54.6	34.1	20.9	40.0	29.6	33.3	37.5	35.6	33.1

TABLE 33.

PERCENTAGE OF ANSWERS GIVEN TO EACH QUESTION. QUESTIONNAIRE STUDY.

PRAISED GROUP.

Questions.	8th Grade.			5th Grade.			3rd Grade.			Totals.
	Boys.	Girls.	Combined.	Boys.	Girls.	Combined.	Boys.	Girls.	Combined.	
	%	%	%	%	%	%	%	%	%	%
1.										
Yes.	60.0	63.6	61.7	83.3	70.0	77.3	57.1	58.3	57.8	65.4
No.	40.0	36.4	38.3	16.7	30.0	22.7	42.9	41.7	42.2	34.6
2.										
Yes.	64.0	81.8	72.3	62.5	100.0	79.6	80.9	75.0	77.7	76.5
No.	36.0	18.2	27.7	37.5	0	20.4	19.1	25.0	22.3	23.5
3.										
Yes.	100.0	100.0	100.0	87.5	100.0	93.2	95.2	91.6	93.3	96.3
No.	0	0	0	12.5	0	6.8	4.8	8.4	6.7	3.7
4.										
Yes.	100.0	100.0	100.0	95.8	95.0	95.4	100.0	87.5	93.3	94.8
No.	0	0	0	4.2	5.0	4.6	0	12.5	6.7	5.2
5.										
First.	12.0	9.1	10.6	33.3	20.0	27.3	57.1	16.6	35.5	24.2
Last.	0	0	0	16.7	5.0	11.3	0	4.1	2.2	4.4
Thruout.	88.0	90.9	89.4	50.0	75.0	61.4	42.9	79.3	62.3	71.4
6.										
Yes.	100.0	100.0	100.0	91.6	100.0	95.4	100.0	100.0	100.0	98.5
No.	0	0	0	8.4	0	4.6	0	0	0	1.5

TABLE 34.

PERCENTAGE OF ANSWERS GIVEN TO EACH QUESTION. QUESTIONNAIRE STUDY.

REPROVED GROUP.

Questions.	8th Grade.			5th Grade.			3rd Grade.			Totals.
	Boys.	Girls.	Combined.	Boys.	Girls.	Combined.	Boys.	Girls.	Combined.	
	%	%	%	%	%	%	%	%	%	%
1.										
Yes.	80.0	54.5	68.1	41.7	45.0	43.1	47.6	29.1	37.7	50.0
No.	20.0	45.5	31.9	58.3	55.0	56.9	52.4	70.9	62.3	50.0
2.										
Yes.	88.0	54.5	72.3	58.3	50.0	54.5	33.3	45.8	40.0	55.8
No.	12.0	45.5	27.7	41.7	50.0	45.5	66.7	54.2	60.0	44.2
3.										
Yes.	100.0	100.0	100.0	100.0	100.0	100.0	90.5	91.6	91.1	97.1
No.	0	0	0	0	0	0	9.5	8.4	8.9	2.9
4.										
Yes.	100.0	100.0	100.0	100.0	100.0	100.0	76.1	100.0	88.8	96.3
No.	0	0	0	0	0	0	23.9	0	11.2	3.7
5.										
First.	68.0	50.0	59.6	25.0	60.0	40.9	57.1	58.3	57.7	52.9
Last.	4.0	4.5	4.2	25.0	10.0	18.1	14.3	8.3	11.1	11.0
Thruout.	28.0	45.5	36.2	50.0	30.0	41.0	28.6	33.4	31.2	36.1
6.										
Yes.	92.0	40.9	68.1	75.0	40.0	59.1	85.7	66.7	75.5	67.6
No.	8.0	59.1	31.9	25.0	60.0	40.9	14.3	33.3	24.5	32.4

In Tables 32, 33 and 34 are shown the percentage of children giving each answer. Table 32 shows the results from the questionnaire given to the Control Group, Table 33, the Praised Group; and Table 34, the Reproved Group. The answers for the three grades are separated, as well as those for boys and girls. Combined results for each grade and for the three grades within each experimental group are also given.

No attempt will be made to analyze the results for the different grades, but the percentage of combined answers for each of the experimental groups will be compared with those of the other two.

Question 1 on all of the Questionnaires, "When you finished the test last week, did you think that you had made a high score on it?", was answered in the affirmative by 60% of the Control Group, 65% of the Praised, and 50% of the Reproved. Suggestion, resulting from the examiner's criticism of the work, seems to have been operative here. Those who were told that they did well in the test, tended to answer in the affirmative more often than those who were told that they did badly.

Question 2 for Control Group, and 3 for Praised and Reproved Groups, "Do you think that you did better in this test

than in the last one?", was answered "Yes" by 79% of the Control Group, 96% of the Praised Group, and 97% of the Reproved Group. It seems as if the children, as a whole, thought they did better after being praised and reproved than after merely repeating the test. It might be interesting at this point to contrast the results of the Questionnaires with the results objectively obtained from the tests. The results of the tests, in terms of percentages of children improving their scores in the second tests, were found to be, 52% for the Control, 79% for the Praised, and 80% for the Reproved Groups. Although these percentages are decidedly lower than in the case of the introspective reports, the order is the same, and the differences between the three groups are approximately the same.

Question 2 of the Praised and Reproved Groups, "Were you surprised when I told you that you had been selected from the whole group because of your very good (poor) work in the last test?", showed rather unexpected results. Seventy-seven per cent of the Praised Group confessed their surprise at being chosen because of excellent work and 56% of the Reproved Group for their poor work. Now, in the latter case, the results are about as one might expect. A certain percentage of these children were accustomed to being failed and this was not an unusual reprimand. The rest of that group, perhaps, felt that because they had not answered all the questions on each page, they too had failed. But, in the case of the praised children, we should have expected only a small percentage, those who usually stood at the foot of the class, to show surprise. A sense of modesty may have been responsible for this.

Question 3 of the Control and 4 on the other two Questionnaires, "Did you try harder in the second test?", showed a true difference between the groups. Fifty-seven per cent of the Control Group, 95% of the Praised Group, and 96% of the Reproved Group said that they did try harder. Although the percentage of children actually improving was not so great, almost every child, after praise or reproof, tried harder. This looks as if the children really felt that the examiner's comments, which preceded the test, were an incentive to work harder.

Question 4 for the Control Group, and 6 for the Praised and Reproved Groups, seemed to have the most significant results. This question, "Did you enjoy taking this test more

than the last one?", was answered "yes" by 67% of the Control Group, 99% of the Praised, and only 68% of the Reproved. Although, according to their own reports, the children, after both praise and reproof, tried harder in the second test than in the first, and although the results actually obtained were practically equal for the two groups, the same feeling of enjoyment was not present. Pride, doubtless, caused the Reproved Group to work harder and do better, but their attitude towards the test was not the same as if praise had been used. Interest and enjoyment in their work is not fostered by any incentive no matter how effective it may appear to be, if a pleasurable feeling tone does not accompany the work.

Question 5 for the Praised and Reproved Groups, "How long were you encouraged (or discouraged) by what I said?" showed that encouragement had a more lasting effect than had discouragement. Seventy-one per cent of the Praised Group said that encouragement lasted throughout the test, while only 36% of the Reproved Group were influenced thus. On the other hand, 53% of the children who had been reproved were discouraged only at the beginning of the second test, and only 24% of the encouraged ones were influenced for so short a time.

Summary

1. According to the children's answers, we may see that they thought they did better work after praise and reproof than after mere repetition of the test.
2. The children in both the Praised and Reproved Groups tried harder in the second test than did the children of the Control Group.
3. The children said that they enjoyed taking the second test more when praised than when reproved, or when no comment was made.
4. Encouragement had a more lasting effect on the group than had discouragement.

9. *Correlations Between First and Second Tests*

As a means of determining the relationship between the relative ranks within the groups of each individual, and the degree to which the children changed their positions in the

second test, correlations between the results of the two tests were computed. The formula used was the Spearman Rank Formula for Correlation,

$$\rho = 1 - \frac{6\Sigma D^2}{n(N^2 - 1)}$$

Now, if the correlations found are lower for the Praised and Reproved Groups than for the Control Group, we may assume that these factors have had an advantageous effect on some of the children, and a disadvantageous effect on others within the same group.

In Table 35 are presented these correlations, together with the P. E.'s of the correlations. The results for the boys and girls taken separately are shown, as well as the combined results. The girls, as a whole, in the three grades of the three experimental groups, show a higher correlation in every case than do the boys, except the 8th grade girls of the Reproved Group. All of the correlations, except the one mentioned, are from 93 to 98, while the correlations for the boys range from 58 to 89. It seems, therefore, as if the boys change their relative positions in every group more often than do the girls of the same groups.

Considering the groups themselves, the Reproved Group for both boys and girls contains the lowest correlations as well as the highest. The coefficients for all three groups, however, are so nearly alike that the only distinctions seem to be due to sex, rather than to the influence of the external factors present.

TABLE 35.

CORRELATIONS BETWEEN RANKS IN FIRST AND SECOND TESTS.

	Control Group.		*Praised Group.*		*Reproved Group.*	
	ρ	*P.E.*	ρ	*P.E.*	ρ	*P.E.*
Boys.						
8th Grade.	.82	±.04	.89	±.03	.87	±.03
5th Grade.	.89	±.03	.81	±.04	.89	±.03
3rd Grade.	.76	±.06	.79	±.05	.58	±.09
Girls.						
8th Grade.	.94	±.02	.95	±.02	.86	±.03
5th Grade.	.95	±.02	.95	±.02	.96	±.01
3rd Grade.	.93	±.02	.93	±.02	.98	±.005
Combined.						
8th Grade.	.92	±.01	.92	±.01	.89	±.02
5th Grade.	.95	±.01	.89	±.02	.93	±.01
3rd Grade.	.89	±.02	.91	±.01	.91	±.01

The results for boys and girls combined show likewise small differences. We may thus conclude that practice, praise and

reproof have little effect on the relative position of the individuals within the groups, when the two sexes are combined, and that the boys change their positions much more frequently than do the girls, who change theirs hardly at all.

Gates and Rissland[4] in their study of the effect of encouraging and discouraging comments on the work of college students in two simple tests, found negative correlations between the first and second tests. For the co-ordination test, the coefficients were –.33, –.03 and –.34; and for the color-naming test, –.24, .16 and –.20 for the Encouraged, Discouraged, and Repetition Groups respectively. These findings differ greatly from those of the present experiment. In every case, the correlations of this study are positive, and all are above the only positive correlation of Gates' study, which is .16.

Summary

1. The correlations between the first and second tests were found to be positive in every case.

2. The correlations were approximately the same for all three groups, showing that external factors had little effect on the relative positions of the subjects within the groups.

3. In the case of the boys, the correlations were much lower than for the girls of the same groups. This shows that the boys changed their positions within the groups more frequently than did the girls and would be explained if the girls, in the initial test, worked more nearly up to their ability than the boys, thus the boys would have a greater margin for improvement due to incentives.

SECTION IV

Summary and Conclusions

The object of this study, as was stated in the Introduction, was to determine the value of praise and reproof as incentives for children. The results have been considered from the point of view of every variable factor entering into the conditions of the experiment, with an attempt to evaluate these incentives under different conditions. At the end of each of the separate sections of results, a brief summary has been given, so that here only the most essential and outstanding of those results will be listed.

The results of the experiment seem to justify the following conclusions:

1. That praise and reproof are incentives which may be used effectively as motivation for school work, and that on the whole, they are of equal value.

2. That older children respond more to both praise and reproof than do younger ones. Praise is slightly more effective for older children and reproof for younger.

3. That boys do better following both praise and reproof than do girls of similar ages. Praise was found to bring about somewhat better work in the case of the girls, and reproof in the case of the boys.

4. That some incentive is more essential for "superior" children than for "inferior" children if their work is to be kept up to the maximum of their ability. Both praise and reproof seemed to serve this purpose. The "superior" children were greatly influenced by both incentives while the "inferior" were decidedly less so. Of the two, praise was somewhat better for the "inferior" and reproof for the "superior" children.

5. That reproof had its greatest effect on children rated by their teachers as "superior" in school work, while praise was of greater value for the "average" and "inferior" children.

6. That negro children react more favorably to praise and white children to reproof. The differences between the groups, however, were only slight, and show that the two races are more alike than different in respect to their reactions to these incentives.

7. That a study of the questionnaires answered by the children at the end of the experiment throws some light on the children's reactions to the incentives. According to their answers, it seems that they tried harder after both praise and reproof; that they thought that they did better in the second tests; but that they enjoyed taking the test more after being praised than after being reproved. In addition to these, they said that encouragement had a more lasting effect than had discouragement.

8. That the correlations between the tests showed that the relative positions within the groups changed less in the case of the girls than in the case of the boys. The differences in the correlations of the three experimental groups were too small to be of significant value.

Although the results indicate that, on the whole, praise and reproof are approximately equal in value as incentives for children, we cannot say that this is always the case. It so happened that the situation of this experiment was such that a certain amount of fear was aroused, following reproof for failure. The children were afraid that they would be dropped into lower grade sections, and this acted as an additional incentive. Especially in the case of the older children was this true. The 8th grade children seemed to realize that failure meant the possibility of not being promoted, and the seriousness of the situation appealed more forcibly to them than to the younger ones.

The pedagogical significance of this study may be touched upon briefly in conclusion. Since boys, on the whole, need an incentive to bring their work up to maximum more often than do girls of the same ages, it would seem that one of the greatest problems of a school would be to supply incentives of different sorts, so as to obtain the best results from the boys. From a comparison of the work done by the different grades, following praise and reproof, it appears that boys and girls of the younger ages respond less to these incentives than do the more mature children. This suggests that other incentives be substituted in the lower grades, which are more suited to the development of the children.

There is no question about the fact that in the ordinary class room it is the dull children who are constantly subjected to reproof for poor work, while the brighter children receive what little praise is given. From the recent work of Pintner

and others, it is now clearly demonstrated that it is the "inferior" children who are working up to capacity, while the "superior" are working at a much lower level than their innate ability would permit. Hence, the inference is that the dull children of a class be encouraged whenever possible, and what reproof is used should be reserved for the most intelligent children who are satisfied with only average work.

From the results of the tests themselves, one might conclude that both praise and reproof are valuable in their effect on children, and that it makes little difference which is used, since the results obtained are so similar in both cases. But a study of the answers given by the children in the Questionnaires shows that the children's attitudes towards the two are not the same. Although they say they try harder regardless of the incentive, they also say that they enjoyed the test more after an encouraging comment than after a discouraging one. Since the two, as we have said, bring about similar results, would it not be better to substitute praise for reproof whenever possible, with the hope of fostering the spirit of enjoyment of and interest in school work, in place of the lack of interest which now prevails?

This study has attempted to answer as conclusively as possible several questions in connection with the use of praise and reproof as incentives for children, and the relative values of the two. There are many others which are suggested from the results obtained, and which might give further information about the subject. The most important of these, it seems to the writer, is to determine whether or not reproof is not to a certain extent conditioned by the child's previous experience and training. Would children, for instance, who were from a distinctly superior environment, react in the same way as these children did, who came from an average or slightly inferior type of environment? Might they not, as a whole, do worse as the result of reproof, rather than better? Another question of almost equal importance, is in connection with the matter of continued reproof, and its effect on the child's work. Kirby, Chapman and Feder, and Book, all of whom have used the practice experiment technique, have found that encouragement, continued over a period of time, proved to be of extraordinary value as an incentive for learning. The study reported in the preceding pages has shown that reproof is as effective an incentive as praise, but, if continually used, might

it not show reverse effects? These and other questions of similar nature will have to be experimentally studied before one can say with absolute assurance just how valuable praise and reproof are as incentives for children.

BIBLIOGRAPHY.

1. George F. Arps: Work With Knowledge of Results vs. Work Without Knowledge of Results. Pysch. Monog. 1920, XXVIII, No. 3.
2. Wm. F. Book and Lee Norvell: The Will to Learn. Ped. Sem. 1922, XXIX, pp. 305-362.
3. J. Crosby Chapman and R. B. Feder: The Effect of External Incentives on Improvement. J. of Educ. Psych. 1917, 8, pp. 469-474.
4. G. S. Gates and L. Q. Rissland: The Effect of Encouragement and of Discouragement upon Performance. J. of Educ. Psych. 1923, XIV, pp. 21-27.
5. Edward P. Gilchrist: The Extent to Which Praise and Reproof Affect a Pupil's Work. School and Society. 1916, IV, pp. 872-874.
6. Walter S. Hunter: General Psychology. U. of Chicago Press, Chicaga, Ill. 1923. pp. 17-39.
7. Albert M. Johanson: The Influence of Incentive and Punishment upon Reaction time. Archives of Psych. 1922, 54.
8. Charles H. Judd: Practice Without Knowledge of Results. Psych. Review Mon. Supplements, 1905, 7, pp. 185-199.
9. Thomas J. Kirby: Practice in the Case of School Children. C.U. Contributions to Educ. T.C. Series. No. 58.
10. F. B. Knight and H. H. Remmers: Fluctuations in Mental Production when Motivation is the Main Variable. J. of App. Psych. 1923, VII, pp. 209-224.
11. Donald A. Laird: How the High School Student Responds to Different Incentives to Work. Ped. Sem. 1923, pp. 358-365.
12. Donald A. Laird: How the College Student Responds to Different Incentives to Work. Ped. Sem. 1923, XXX, pp. 366-370.
13. William McDougall: Social Psychology. Luce and Co., Boston, 1921, p. 186.
14. J. H. McFadden and J. Dashiell: Racial Differences as Measured by the Downey Will-Temperament Tests. J. of App. Psych. 1923, 7, pp. 30-53.
15. Norsworthy and Whitley: Psychology of Childhood. Macmillan and Co. 1923, p. 109.
16. R. Pintner and H. Marshall: A Combined Mental-Educational Survey. J. of Educ. Psych. 1921. XII, pp. 32-43.
17. Albertine A. Richards: Motive in Education. Ped. Sem. 1921, XXVIII, pp. 60-72.
18. Rietta Simmons: The Relative Effectiveness of Certain Incentives in Animal Learning. Comp. Psych. Monog. 1924. II, No. 7.
19. Llewellyn T. Spencer: The Effects of Practice Without Knowledge of Results. Am. J. of Psych. 1923. vol. 34, pp. 107-111.
20. C. W. Stone and Carl Colvin: "How to Study" as a Source of Motive in Educational Pyschology. J. of Educ. Psych. 1920, Vol. 11, pp. 348-354.
21. L. M. Terman: The Measurement of Intelligence. Houghton Mifflin Co. 1916. p. 79.
22. Herbert Toops and P. M. Symonds: What Shall We Expect of the A. Q.? J. of Educ. Psych. 1922, XIII, p. 513.
23. Margaret F. Washburn: The Animal Mind. Macmillan Co. 1917, pp. 258-278.
24. Wm. R. Wright: Some Effects of Incentives on Work and Fatigue. Psych. Review. 1906, vol. 13, pp. 23-34.

VITA.

The writer of this dissertation was born in Harrisburg, Penna., on July 4, 1898. She received her college preparatory training at the Seiler School, of Harrisburg, from which she was graduated in June, 1915. In 1919 she received the degree of Bachelor of Arts from Bryn Mawr College, and in 1922, the degree of Master of Arts, with Psychology as a Major, and Education and Philosophy as a Minor subject. During the year 1919-1920, she taught in the Manor School, at Larchmont Manor, N. Y., and from 1920-1921, was head of the Department of Mathematics at the Springside School, of Chestnut Hill, Penna. From 1922-1924, she was a graduate student in the Department of Psychology of Columbia University, and received the degree of Doctor of Philosophy from that University in June, 1924.